IDENTIFICATION AND CHARACTERIZATION OF DIATOM KINASES CATALYZING THE PHOSPHORYLATION OF BIOMINERAL FORMING PROTEINS

A Dissertation
Presented to
The Academic Faculty

by

Vonda Chantal Sheppard

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Chemistry and Biochemistry

Georgia Institute of Technology
December 2010

IDENTIFICATION AND CHARACTERIZATION OF DIATOM KINASES CATALYZING THE PHOSPHORYLATION OF BIOMINERAL FORMING PROTEINS

Approved by:

Dr. Nils Kröger, Advisor
School of Chemistry and Biochemistry
School of Materials Science and Engineering
Georgia Institute of Technology

Dr. Valeria Milam
School of Materials Science and Engineering
Georgia Institute of Technology

Dr. Nicholas Hud
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Raquel Lieberman
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Christine Payne
School of Chemistry and Biochemistry
Georgia Institute of Technology

Date Approved: November 12, 2010

ACKNOWLEDGMENTS

I would like to express my deepest appreciation for my advisor Dr. Nils Kröger whose great scientific knowledge, enthusiasm, and patience made this thesis possible. His dedication to science has made him as a constant source of ideas which exceptionally inspired my growth as a researcher and a scientist. Above all and the most needed, he provided me unflinching encouragement and support in various ways. It is quite difficult to convey the magnitude of my appreciation.

I would like to give Dr. Nicole Poulsen a great big THANK YOU! She is truly amazing and if it were not for her, the lab would not function as it does. Her willingness to help and her compassion for people is truly amazing. I owe her as much gratitude as I do Dr. Kröger.

I would also like to thank all of my committee members, Dr. Nicholas Hud, Dr. Christine Payne, Dr. Valeria Milam, and Dr. Raquel Lieberman for guiding me to be a better scientist and for your valuable input during our meetings. A special thanks to Dr. Bridgette Barry for her wisdom and encouragement.

Thanks to all of my co-researchers (past and present) for helping me get through what has seemed like an endless journey to obtain a doctorate. Your great ideas and the social time we spent outside of lab really got me through some tough times.

Last, but certainly not least, I would like to thank all of my family and friends for their support and love throughout this process. Without you, I do not know where I would be right now. I owe you guys so much.

This work was supported by the Georgia Institute of Technology, the National Science Foundation (DMR# 0845939), the Research Corporation for Science Advancement, and the U. S. Department of Education.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF COMMONLY USED BUFFERS

1x PCR Buffer	50 mM Tris-HCl pH 8.8, 50 mM NaCl, 2.5 mM $MgCl_2$, 2 mM DTT)
TE	10 mM Tris-HCl pH 8.0, 1 mM EDTA
ECLB	50 mM Tris-HCl pH 8.0, 1 M NaCl, 1 mM PMSF, 5 mM Imidazole
TBS	25 mM Tris-HCl pH 7.5, 100 mM NaCl
MB	50 mM HEPES-NaOH pH 7.5, 150 mM NaCl, 1 mM EDTA, 0.2 mM DTT, 10 mM NaF, 1 mM Na_3VO_4
PBS	20 mM sodium phosphate pH 7.5, 150 mM NaCl

LIST OF SYMBOLS AND ABBREVIATIONS

ATP	adenosine triphosphate
NAD	nicotinamide adenine dinucleotide
GTP	guanosine triphosphate
EDTA	ethylenediaminetetraacetic acid
PCR	polymerase chain reaction
Ci	curie
IDP	inositol diphosphate
n	nano
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
min	minute
hr	hour
Ex.	excitation
Em.	emission
RNA	ribonucleic acid
DNA	deoxyribonucleic acid
dNTPs	deoxynucleoside triphosphates
poly A	polyadenosine
poly C	polycytosine
Ni-NTA	Ni^{2+}-nitriloacetic acid
U	units
m	milli
sec	second
Tris	tris(hydroxymethyl) aminomethane
μ	micro
GFP	green fluorescent protein

LIST OF VENDORS

5 PRIME (Gaithersburg, MD)

Acros Organics (Morris Plains, NJ)

Agilent (Santa Clara, CA)

BioRAD (Hercules, CA)

Biotek (Winooski, VT)

Calbiochem (Gibbstown, NJ)

ClonTech (Mountain View, CA)

Corning (Corning, NY)

EMD Biosciences (Gibbstown, NJ)

Fermentas (Glen Burnie, MD)

Fujifilm (Stamford, CT)

GE Healthcare (Piscataway, NJ)

Genescript (Piscataway, NJ)

Invitrogen (Carlsbad, CA)

Kodak (Rochester, NY)

Lampire Biological Laboratories (Pipersville, PA)

Life Technologies (Carlsbad, CA)

Machery-Nagel (Düren, Germany)

Millipore (Billerica, MA)

Misonix (Farmingdale, NY)

MP Biomedicals (Solon, OH)

Novagen (Carlsbad, CA)

Pierce (Rockford, IL)

Promega (Madison, WI)

Qiagen (Valencia, CA)

Roche (Florence, SC)

Sartorius (Germany)

Shimadzu (Columbia, MD)

Sigma-Aldrich (St. Louis, MO)

Whatman (Piscataway, NJ)

Zeiss (Germany)

SUMMARY

Diatoms are unicellular photosynthetic algae that display intricately patterned cell walls made of amorphous silicon dioxide (silica). Long-chain polyamines and highly phosphorylated proteins, silaffins and silacidins, are believed to play an important role in biosilica formation. The phosphate moieties on silaffins and silacidins play a significant role in biomineral formation, yet no kinase has been identified that phosphorylates these biomineral forming proteins. This dissertation describes the characterization of a novel kinase from the diatom *Thalassiosira pseudonana*, tpSTK1, which is upregulated during silica formation. A recombinantly expressed histidine-tagged version of tpSTK1 was capable of phosphorylating recombinant silaffins but not recombinant silacidin *in vitro*. Through establishing methods for subcellular fraction of *T. pseudonana* membranes in combination with antibody inhibition assay, it was discovered that native tpSTK1 phosphorylates silaffins but not silacidins *in vitro* (i.e. it exhibits the same substrate specificity as recombinant tpSTK1). As tpSTK1 is an abundant protein in the ER lumen (~ 0.5 % of total ER protein) it seems highly likely to function as a silaffin kinase *in vivo*. TpSTK1 lacks clear sequence homologs in non-diatom organisms and is the first molecularly characterized kinase that appears to be involved in biomineralization.

The predicted kinase domain (KD) of tpSTK2, the only *T. pseudonana* homolog of tpSTK1, was recombinantly expressed and tested for phosphorylation activity. Recombinant tpSTK2-KD and native tpSTK2 exhibited detectable activity with myelin basic protein, but did not phosphorylate silaffins or silacidins *in vitro*. Western blot analysis demonstrated that native tpSTK2 was not present in the ER, but associated with the cytosol and Golgi membrane containing subcellular fractions.

CHAPTER 1: INTRODUCTION

1.1 Overview of Biomineralization

Understanding how living organisms synthesize inorganic substances mediated by organic molecules in a process called biomineralization has intrigued scientists for hundreds of years. A wide range of organisms, from bacteria to humans, produce a variety of inorganic materials such as iron oxide, calcium carbonates, calcium phosphates, and silicates. The "biocomposites" (the term for the combination of organic material and minerals[1]) formed by these organisms exhibit complex, species-specific morphologies strikingly different from their inorganically formed counterparts (Figure 1).

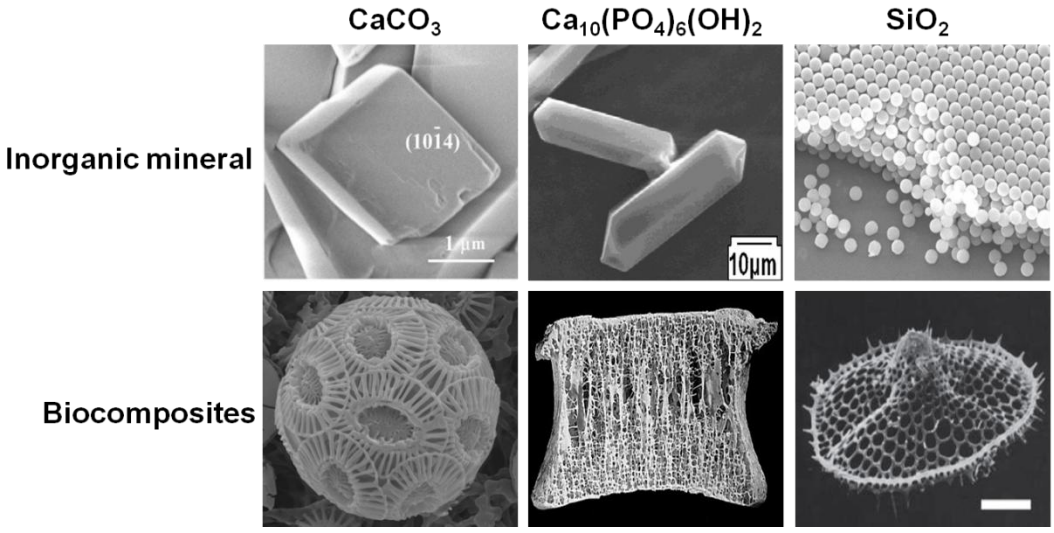


Figure 1.1. Comparison of mineral structures. SEM images of ***Top***, inorganically formed calcite ($CaCO_3$)[2], hydroxyapatite $Ca_{10}(PO_4)_6(OH)_2$[3], and silicon dioxide (SiO_2)[4] crystals. ***Bottom***, Biocomposite formation of cococoliths by the phytoplankton *Emiliania huxleyi* [5] (left), bones in humans[6] (center), and cell walls in a radiolaria[7] (right) using the same minerals, respectively.

The biologically mediated formation of these highly ordered, intricately structured biocomposites at ambient conditions is of great interest for materials scientists, who are often unable to synthesize complex chemical structures. Synthetic reproduction of these biocomposites would require identification of the key components involved in each stage of biomineralization: nucleation, organic matrix formation, and crystal growth regulation[8]. Although the biocomposites vary in their chemical composition, structures, and function, there are some similarities in their biomineralization processes. In all of the organisms studied to date, biomineral formation and morphogenesis occurs within specialized extra- or intracellular compartments mediated by a matrix of macromolecules that includes proteins and polysaccharides[1,8]. Specifically, acidic phosphoproteins have been identified in organic matrices from a diverse set of biominerals (from marine algae to humans) and that may interact with mineral ions or the structural framework of the matrices[1,8-10]. In the following, several well characterized phosphoproteins involved in the formation of the most abundant biominerals (calcium carbonate, hydroxyapatite, and silica) will be discussed.

1.2 Phosphoproteins Involved in Biomineralization

1.2.1 Calcium Carbonate and Calcium Phosphate Biomineral Proteins

Comprising nearly 50 % of all known biominerals, calcium-containing minerals are the most abundant biominerals on Earth[8-9]. Due to the predominance of calcium-containing biominerals, the field of biomineralization used to be known as “calcification” until the early 1980s as a more and more biominerals were being discovered[8-9]. Calcium

phosphate in the form of hydroxyapatite ($Ca_{10}(PO_4)_6(OH)_2$) and calcium carbonate ($CaCO_3$) are the two major calcium containing biominerals[1,8-10].

1.2.1.1 Hydroxyapatite formation

Bone and dentin are mineralized tissues found in vertebrates that contain hydroxyapatite as the mineral component. The mineralization of bone (osteogenesis) and teeth (dentiogenesis) is a highly controlled process that occurs by the deposition of hydroxyapatite crystals ($Ca_{10}(PO_4)_6(OH)_2$) in an extracellular organic matrix consisting of type I collagen and several non-collagenous proteins (NCPs)[12]. Collagen serves as the structural framework of the organic matrices from bone and teeth and its mineralization is mediated by NCPs that are believed to induce and control hydroxyapatite formation[12]. Among the NCPs are the bone matrix proteins osteopontin (OP) and bone sialoprotein (BSP), and the dentin matrix protein phosphophoryn (PP) that belong to the SIBLING family (small integrin binding ligand *N*-linked glycoprotein). Although there is little sequence homology amongst the proteins in this family, they do share similar features: 1) presence of an Arg-Gly-Asp (RGD) motif that mediates cell attachment via its interaction with integrins, 2) prevalence of acidic amino acids (i.e. glutamate and/or aspartate), and 3) modifications by posttranslational processing[12]. It has been demonstrated that the activities of OPN, BSP, and PP are influenced by their degree of phosphorylation.

In the process of bone remodeling, mature bone is removed (bone resorption) by osteoclasts and replaced with new bone (ossification) by osteoblasts. During bone remodeling, bone cells deposit various hypo-phosphorylated forms of the bone matrix proteins OP and BSP (ranging from 1 to 14 nmol phosphoserine/mol protein[13])

suggesting that the degree of phosphorylation of these proteins is important in bone remodeling[14-18]. It has been demonstrated *in vitro* that the binding of bone cells to OPN and BSP (primarily achieved through the interactions of bone cell surface integrins with an RGD motif) can be enhanced by 40 % depending on the degree of phosphorylation of OPN and BSP[19-20]. In addition to bone cell attachment, BSP and OPN also regulate the bone mineralization process by acting as a nucleator or inhibitor, respectively, of hydroxyapatite formation. These activities are completely dependent on the phosphorylation level of the bone matrix proteins[21-26]. Knockout mice lacking OPN or BSP resulted in excessive mineralization of bone or stunted and demineralized bone, respectively, which is consistent with the inhibition and nucleation activity of these proteins *in vitro*[27-28].

OPN and BSP are also found in dentin but not as the predominant proteins[29]. Instead, dentin sialoprotein (DSP) and dentin phosphoprotein (also known as phosphophoryn, PP) and are the major NCPs found in dentin[12,29-31]. PP is the most abundant protein in the dentin extracellular organic matrix and is likely a key regulator of hydroxyapatite formation[12,29,31]. Mainly composed of aspartate and phosphoserine (~90 % of all residues), phosphophoryn (PP) is a highly acidic phosphoprotein with a strong affinity for calcium ions[12,25,31-34]. The potency of calcium ion sequestering by PP has been attributed to the presence of phosphoserine as phosphorylated PP binds at least 2 fold more calcium than recombinant PP *in vitro*[32]. The phosphate moieties on PP are also responsible for the binding of PP to collagen fibrils, and the nucleation of hydroxyapatite crystals onto these fibrils as dephosphorylation of PP completely abolishes these activities *in vitro*[33-34]. The exact role of DSP is unknown. DSP and PP originate from the

same inactive precursor protein, dentin sialophosphoprotein (DSPP), which is secreted by odontoblasts (tooth cells) and proteolytically processed to yield the two dentin NCPs[35]. Mutations in the *Dspp* gene or knockdown of the gene results in dentiogenesis imperfecta in humans or defective mineralization in mice, respectively, indicating that DSP and PP, are essential for normal hydroxyapatite mineralization in teeth[36-37].

1.2.1.2 Calcium carbonate formation

The most abundant biomineral produced is calcium carbonate ($CaCO_3$), which is synthesized by a very large and diverse set of organisms[1,8-11]. The calcareous structures produced by these organisms primarily serve as an exoskeleton for protection or for the storage of calcium in terrestrial species where calcium availability is limited[11]. Depending on the specific function, these organisms are capable of utilizing one or more of the eight known polymorphs of $CaCO_3$ for biomineralization, with the majority of species favoring calcite, aragonite, and amorphous $CaCO_3$[1,8-11].

As with hydroxyapatite mineralization, an organic matrix (consisting mainly of chitin) is secreted during calcium carbonate formation that is believed to control mineral formation and influence the crystal phase[8-9,11]. The organic matrix is typically divided into two classes based on its solubility after decalcification with EDTA or a weak acid: the soluble matrix and the insoluble matrix[8-9,11]. Generally, the soluble matrix is believed to interact with the mineral phase and is composed mainly of acidic amino acids (i.e. glutamate and aspartate), glycine, and phosphoserine that are believed to interact with the mineral[38-39]. In contrast, the insoluble matrix is thought to form the structural framework of the organic matrix and is generally dominated by the presence of glycine, alanine, and

hydrophobic amino acids[38].

Several soluble matrix proteins, specifically phosphoproteins, from some of the most highly studied calcium carbonate forming organisms (mollusks, crustaceans, and the sea urchin) have been characterized and are likely involved in calcium carbonate biomineralization[40-47]. Terrestrial crustaceans such as *Orchestria cavimana* and *Cherax quadricarinatus* store calcium (originating from the old calcified exoskeleton) in organs known as gastroliths as "concretions" composed of amorphous calcium carbonate precipitated within an organic matrix[40-41]. Two highly acidic phosphoproteins, orchestin and gastrolith protein-10 (GAP-10), identified in the organic matrices of the crustaceans' concretions exhibit calcium binding activity *in vitro*[40-41]. For orchestin, it has been shown that phosphorylation of serine residues is required for its calcium binding activity[41]. The importance of protein phosphorylation of some soluble matrix proteins from molluscan and crustacean exoskeletons has also been demonstrated. Two highly acidic phosphoproteins from mollusk shells (RP-1; from the scallops *Adamussium colbecki* and the oyster *Crassostrea virginica* and molluscan phosphorylated-protein (MPP-1); from the oyster *Crassostrea nippona*), and phosphoprotein calcification-associated peptide-1 (CAP-1) from the cuticle (term for crustacean exoskeleton) of the crayfish *Procambarus clarkii* all inhibit calcium carbonate precipitation *in vitro* due to their phosphate moieties[42-45]. Although the majority of these phosphoproteins from the soluble organic matrix play an inhibitor role in calcium carbonate crystal formation, it has been shown that MPP-1, when premixed with insoluble organic matrix material, can induce crystal formation *in vitro*[45]. Therefore, it is likely that MPP-1 and similar proteins serve a dual function in $CaCO_3$ biomineralization by interacting with the structural framework to

induce mineral formation (analogous to phosphophoryn's role in dentiogenesis). Phosphoproteins UTMP-16 and P19 have also been identified in the $CaCO_3$-based teeth of the sea urchins *Lytechinus variegatus* and *Strongylocentrotus purpuratus*, respectively, and have strong affinities for calcium ions[46-47]. Interestingly, these sea urchin phosphoproteins are so similar to PP that antibodies raised against PP also recognize these proteins[47]. Although the influence of phosphorylation on the calcium binding abilities of UTMP-16 and P19 has not been tested, it is likely (given the similarities to PP) that these phosphoproteins are important proteins in sea urchin tooth formation.

1.2.2 Silica Biomineral proteins

Silicon, the second most abundant element on Earth, is the element of choice (instead of calcium) by organisms such as sponges, radiolarians, and diatoms that produce silicon dioxide (SiO_2, also known as silica) in a process known as "silicification"[1,8,48-51]. Diatoms, the predominant producers of biosilica, are unicellular, photosynthetic algae that account for ~20 % of the Earth's total photosynthetic carbon dioxide fixation[1,8,48]. There are more than 10,000 different species of diatoms in existence each displaying a unique porous, nanopatterned cell walls (termed frustules) made of silicon dioxide[48-51] (Figure 1.2 A). The diatom cell wall consists of two overlapping parts, the epitheca and hypotheca, with each theca containing a valve and several girdle bands[48-51] (Figure 1.2 B).

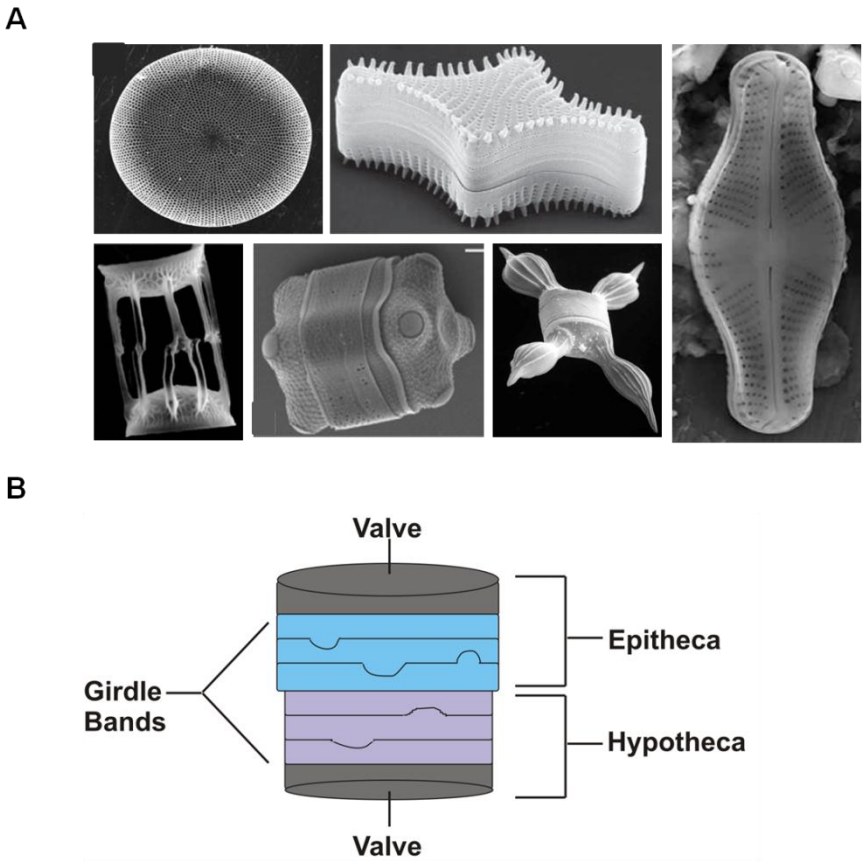


Figure 1.2. Diatoms. ***A,*** SEM images of various species of diatoms[52-57]. Images taken from the web. ***B,*** Schematic representation of the diatom cell wall.

Silica deposition and morphogenesis occurs within a specialized membrane bound compartment termed the silica deposition vesicle (SDV)[48-51]. During cell division and interphase, separate SDVs develop during valve formation and girdle band formation. Subsequently, the complete mature silica structure is exocytosed to the cell surface[48-49,51]. Dissolution of isolated diatom silica cell walls using acidified ammonium fluoride solubilizes proteins (silaffins and silacidins) and long chain polyamines (LCPAs) that are

embedded in the silica[58-63]. Given their intimate association with the diatom silica and their ability to form silica from silicic acid solutions *in vitro*, these proteins and LCPAs are believed to be components of the SDV and involved in diatom silica biogenesis.

1.2.1.1 Silaffins and Silacidins

Silaffins (proteins with silica affinity) are a family of phosphoproteins that possess numerous additional posttranslational modifications[58-61]. Biochemical analysis of the proteolytically derived peptides natSil-$1A_1$, natSil-$1A_2$, and natSil-1B from *Cylindrotheca fusiformis* silaffin Sil1, revealed that all lysine side chains were alkylated and all serine and trimethylhydroxylysine residues are phosphorylated, making it zwitterionic in nature (Figure 1.3)[58-59]. The presence of phosphate moieties greatly influenced the ability of these silaffin-1 peptides to induce silica formation, *in vitro*[58-59]. In the presence of silicic acid, natSil1-A rapidly (within 10 min) precipitates silica in the form of nanospheres but this silica forming activity is completely lost after dephosphorylation of natSil-1A by anhydrous hydrogen fluoride[58-59].

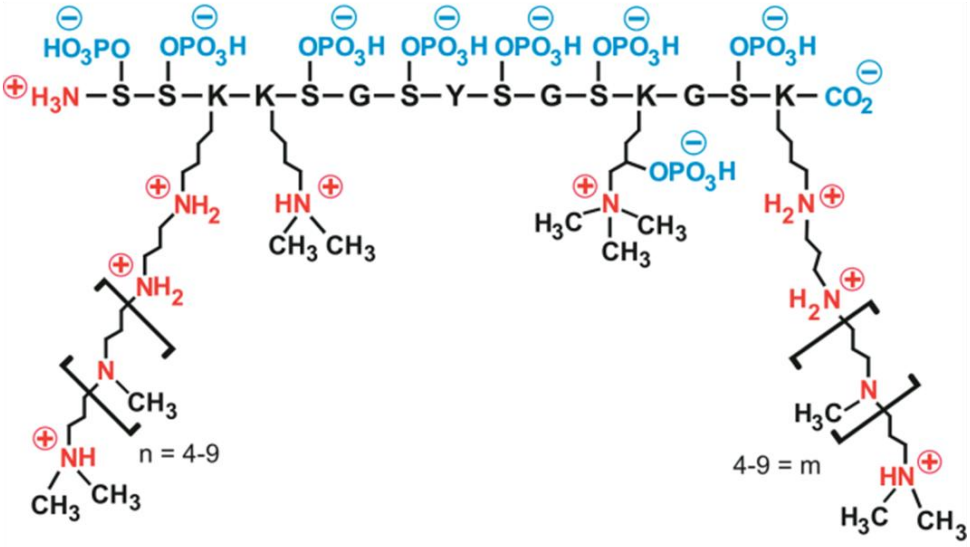


Figure 1.3. Chemical structure of native silaffin-1A$_1$ peptide from the diatom *C. fusiformis*[59]. Image taken reference 59.

Other silaffins from *C. fusiformis* (natSil-2)[60] and *Thalassiosira pseudonana* (tpSil1H, tpSil1L, tpSil2H, tpSil2L, and tpSil3)[61] are not only phosphorylated and contain alkylated lysines and phosphorylation, they are also glycosylated and sulfated. Furthermore, amino acids such as dihydroxyproline (a rare amino acid) and phosphorylated hydroxyproline (a previously unknown amino acid) are present in tpSil1/2H and natSil-2, respectively[60-61]. *T. pseudonana* native silaffins and natSil-2 lack intrinsic silica forming capabilities *in vitro*, but in the presence of LCPA they are able to precipitate silica and control the amount and/or the structure of the silica produced (Figure 1.4 A)[60-61]. Thus, these silaffins are termed regulatory silaffins. Silaffins natSil-2, tpSil1/2H, and tpSil3 (in the presence of constant LCPA concentrations) produced silica structures (i.e. plates and porous silica) that are more reminiscent of diatom biosilica[60-61].

At certain concentrations, however, these silaffins strongly inhibit silica formation (Figure 1.4 B)[60-61]. In contrast, lower molecular weight isoforms of tpSil1/2H (generated by intracellular proteolytic processing), silaffins tpSil1/2L, together with LCPA produce silica spheres that increase in diameter as the silaffin concentration increases[61].

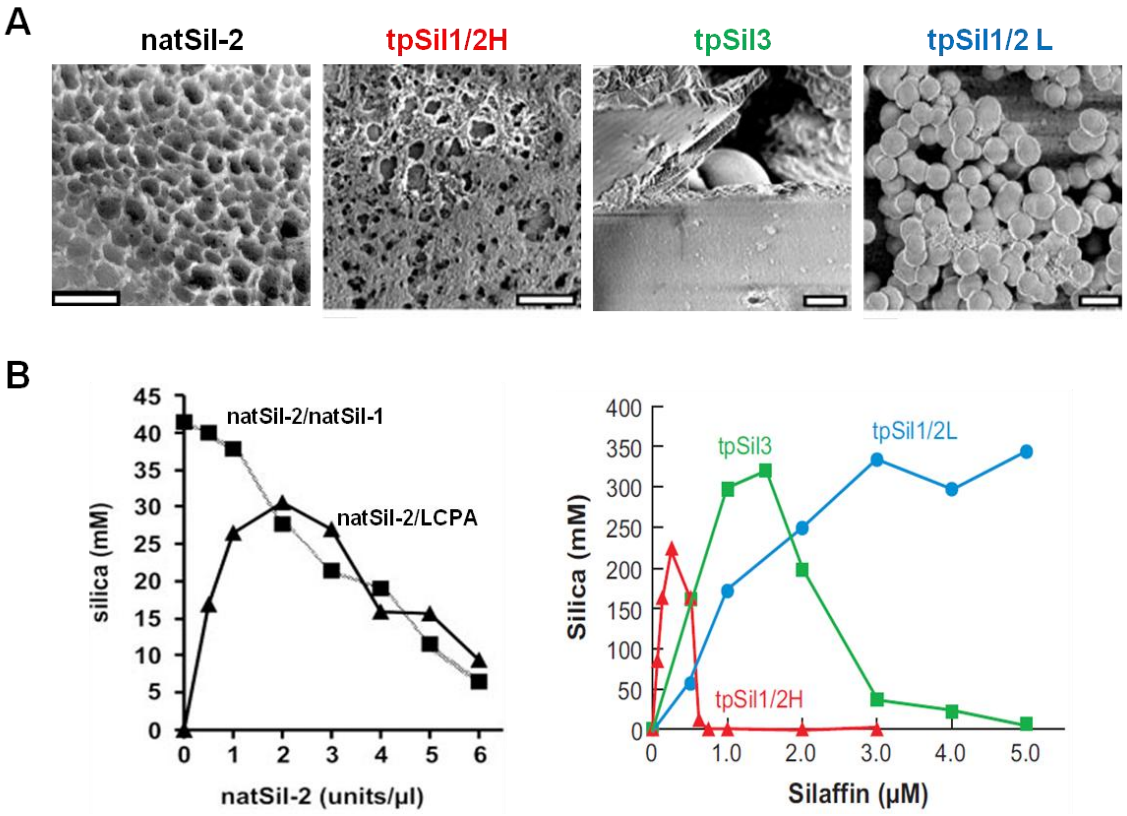


Figure 1.4 Silica precipitating activity of regulatory silaffins from *C. fusiformis* and *T. pseudonana*[60-61]. *A,* SEM images of the silica structure produced by natSil-2/natSil-1 mixtures (black letters) and *T. pseudonana* silaffin-LCPA mixtures (text in red, green, or blue) in the presence of silicic acid. Scale bar for natSil-2 micrograph is 2 µm. All other scale bars are 1 µm. ***B,*** Influence of native silaffins on the silica precipitating activity from the silaffin-LCPA or natSil-2/natSil-1 mixtures. Images taken from references 60 and 61.

The ability of silaffins to inhibit silica formation may be depend on the degree of glycosylation. TpSil1/2L contains about 3-10 times fewer carbohydrates than silaffins tpSil1/2H and tpSil3 and did not inhibit silica formation[61]. At concentrations about one fourth the concentration required for tpSil3-mediated inhibition, tpSil1/2H was a potent inhibitor of silica formation[61]. Interestingly, tpSil1/2H contains nearly four times more carbohydrates than tpSil3 which perfectly correlates (linearly) with the concentration of silaffin required for inhibition[61]. The phosphate moieties on silaffins, on the other hand, appear to only stimulate silica precipitation as tpSil1/L and deglycosylated/desulfated natSil2 (in the presence of LCPAs), both of which consist mainly of phosphate groups, do not inhibit silica formation *in vitro* at any of the silaffin concentrations tested[60-61].

More recently, phosphopeptides termed silacidins were isolated from *T. pseudonana* biosilica[62]. The silacidin peptides (Silacidin A, Silacidin B, Silacidin C) are products of proteolytic processing of a precursor protein rich in serine residues, all of which are phosphorylated, and in acidic amino acids making these peptides the first highly acidic peptides (pI < 3) to be isolated from diatom biosilica[62] (Figure 1.5). There is no evidence of other posttranslational modifications in these peptides. In the presence of LCPA, micomolar amounts of silacidins induced the formation of silica nanospheres whose sizes increased with silacidin concentration and did not, at any concentration tested, inhibit silica formation[62].

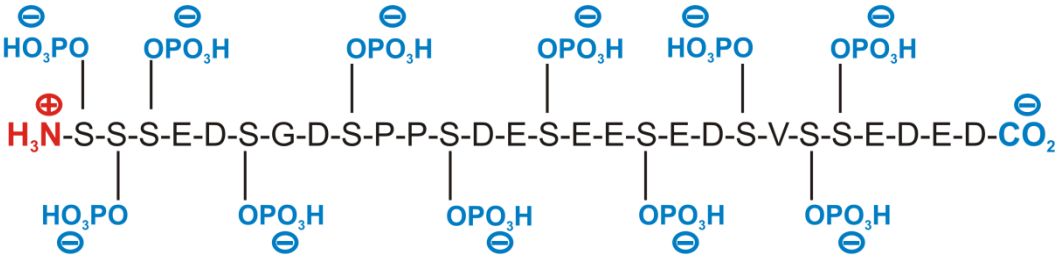


Figure 1.5. Chemical structure of a silacidin peptide from *T. pseudonana*.

1.2.2.2 Long Chain Polyamines

Although long chain polyamines are not phosphoproteins, their intimate association with silaffins and silacidin warranted a discussion on LCPAs and their involvement in diatom biomineralization. All of the LCPAs identified from several species of diatoms have linear oligo-propyleneimine chains that vary in chain length and degree of methylation attached to putrescine, spermidine or propylenediamine[64]. Similar oligo-propyleneimine chains attached to lysine residues are found in silaffins (see Figure 1.3). The tight association of LCPAs with diatom silica and their influence of silica formation in vitro led to the hypothesis that LCPAs are involved in silica formation *in vivo*. Consistent with this hypothesis is the observation by Hildebrand and co-workers that *T. pseudonana* cells treated with 1,3-diaminopropane dihydrochloride, an inhibitor of polyamine biosynthesis, exhibited a marked decrease in silica content of the cell walls and phenotypic abnormalities (Figure 1.6)[65].

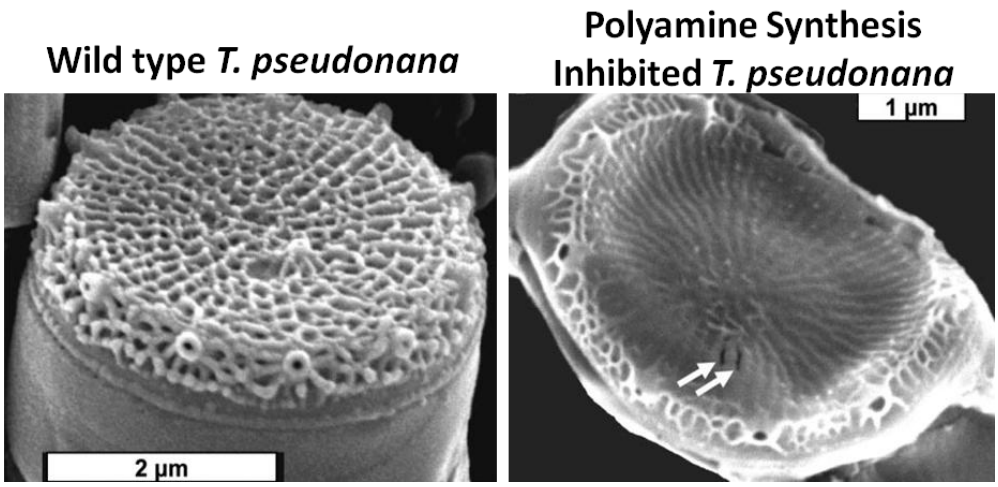


Figure 1.6. Effect of polyamine synthesis inhibition on the biosilica structure of T. pseudonana[65]. SEM micrographs of ***A,*** wild-type cells and ***B,*** cells treated with 10 mM 1,3-diaminopropane dihydrochloride.

In the presence of phosphate (a polyvalent anionic molecule), LCPAs-phosphate microdroplets form resulting in a phase separation from the aqueous solution[66]. The size of the microdroplet is dependent on the ratio of polyamine to phosphate ions and increases as the phosphate concentration increases. Since the microdroplets solidify to form silica spheres upon the addition of silicic acid, the size of the silica spheres formed by the polyamine-phosphate system is also dependent on phosphate concentration[66]. It has been hypothesized that *in vivo*, the highly phosphorylated silaffins and silacidin (instead of inorganic phosphate), are the polyanionic molecules that form electrostatic interactions with the polycationic LCPAs to induce a phase separated nanopatterned organic matrix (dependent on the silaffin) that may act as a template for silica formation[48,66]. The *in vitro* silica formation assays support this hypothesis since biosilica formation only occurred in the presence of LCPAs together with silaffins and/or silacidin

and the structure of the biosilica produce was dependent on the silaffin.

1.3 Regulation of Phosphorylation by Protein Kinases and Phosphatases

The prevalence of phosphoproteins in biomineralizing organisms from many different taxa suggest that they are essential components of biomineral forming machinery. Furthermore, it has been demonstrated that phosphorylation of these proteins is critical for biomineralization activities. So how is the phosphorylation of these biomineralization proteins regulated? Regulation of phosphorylation of biomineral forming proteins is achieved by the action of various protein kinases and phosphatases[67]. In fact, the regulation, function, and intracellular localization of about 30 % of the proteins of a eukaryotic cell, depends on phosphorylation by protein kinases[68]. Protein kinases are enzymes that transfer phosphoryl groups from a donor molecule (usually ATP) in the presence of a divalent metal cation (usually Mg^{2+}) to hydroxyl amino acids (i.e. serine, threonine, and tyrosine) and in rare instances to histidine residues[69, 68] (Figure 1.7). About 80 % of known protein kinases phosphorylate serine or threonine residues and are collectively known as serine/threonine (S/T) kinases[68]. The degree of phosphorylation of a protein can be regulated by phosphatases, which are enzymes that remove phosphate groups in a process known as dephosphorylation. Together, kinases and phosphatases control the number and location of phosphates attached to a protein.

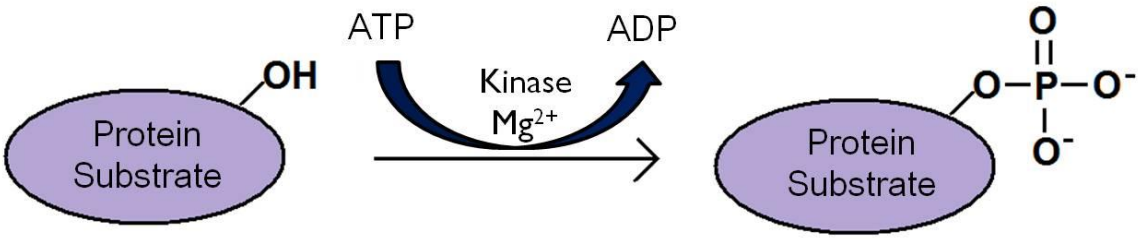


Figure 1.7. Schematic representation of protein phosphorylation.

Tartrate resistant acid phosphatase (TRAP) is an iron-containing enzyme that is secreted by osteoclasts during bone resorption and has been shown to dephosphorylate the bone matrix proteins osteopontin and bone sialoprotein, *in vitro*[19,70]. Mice deficient in active TRAP resulted in mild osteopetrosis (excessive mineralization of bone)[71] and reduced migration of osteoclasts on the bone surface[72]. In contrast, overexpression of TRAP leads to osteoporosis (demineralization of bone)[73]. Reduced phosphorylated forms of OPN and BSP bind less bone cells than their native, fully phosphorylated forms. It is therefore speculated that modulation of the phosphorylation level of bone matrix proteins by TRAP results in a gradient of variant hypo-phosphorylated forms of bone matrix proteins that facilitate the detachment and migration of bone cells along the bone surface and regulate biomineral formation[72].

Although TRAP, a phosphatase, has been shown to be involved in bone biomineralization, not a single kinase has been identified to date that is responsible for phosphorylation of bone matrix proteins. This is quite surprising given that more than 10 years ago *in vitro* phosphorylation of OPN, BSP and dentin protein phosphophoryn (PP) was achieved by yet uncharacterized kinases operating within mammalian ER and Golgi

membrane fractions[74-76] and in a separate study, from kinases present with the extracellular matrix (termed ectokinases[77]) of bone cells[78]. These kinases (denoted as hydroxyapatite biomineralization kinases or HBKs hereafter) specifically phosphorylated serine or threonine residues in the vicinity of acidic amino acids[74-76,78]. Interestingly, the bone and teeth kinases mainly recognized the same consensus sequences, D/E- $X_{1 \text{ or } 2}$-S/T or S/T-X-X-D/E (where X is usually any non-basic amino acid) that are recognized by cytosolic kinases casein kinase I and casein kinase II, respectively[68,74,78]. The HBKs phosphorylated the acidic protein casein and in some instances, could utilized GTP as the phosphate donor, a characteristic of cytosolic CKII[68,74-76,78]. Given the similarities of the HBKs to the cytosolic kinases CKI and CKII, the HBKs are commonly referred to as isoforms of CKI and CKII. However, several kinases (e.g. AMP-activated protein kinase and Polo-like kinase 1) phosphorylate casein[79-81]. In addition, a Golgi-localized HBK in mammalian cells recognized an additional consensus sequence, S-X-E/Sp (Sp is phosphoserine), resulting in phosphorylation of residues that were not phosphorylated by the casein kinases[75]. In the absence of any sequence information on the bone and teeth biomineralization kinases, there exists no definitive evidence that the bone and teeth biomineralization kinases are isoforms of the cytosolic casein kinases.

1.4 Kinases in the Secretory Pathway

Identification and characterization of kinases involved in the phosphorylation of biomineral proteins would not only reveal another component of the biomineralizing machinery, but it would also provide information on kinases operating within the

secretory pathway, which is poorly understood[82-87]. The kinases that regulate phosphorylation of biomineralizing proteins must reside within the ER, Golgi, and/or the specialized biomineral forming compartments. This is due to the fact the biomineral forming proteins (the ones characterized to date) contain an N-terminal signal sequence for co-translational import into the endoplasmic reticulum (ER)[12,48,74,82]. During their transit from the ER to the mineral forming compartment, biomineralization proteins are posttranslationally modified and at no time are they exposed to the cytosol (Figure 1.8).

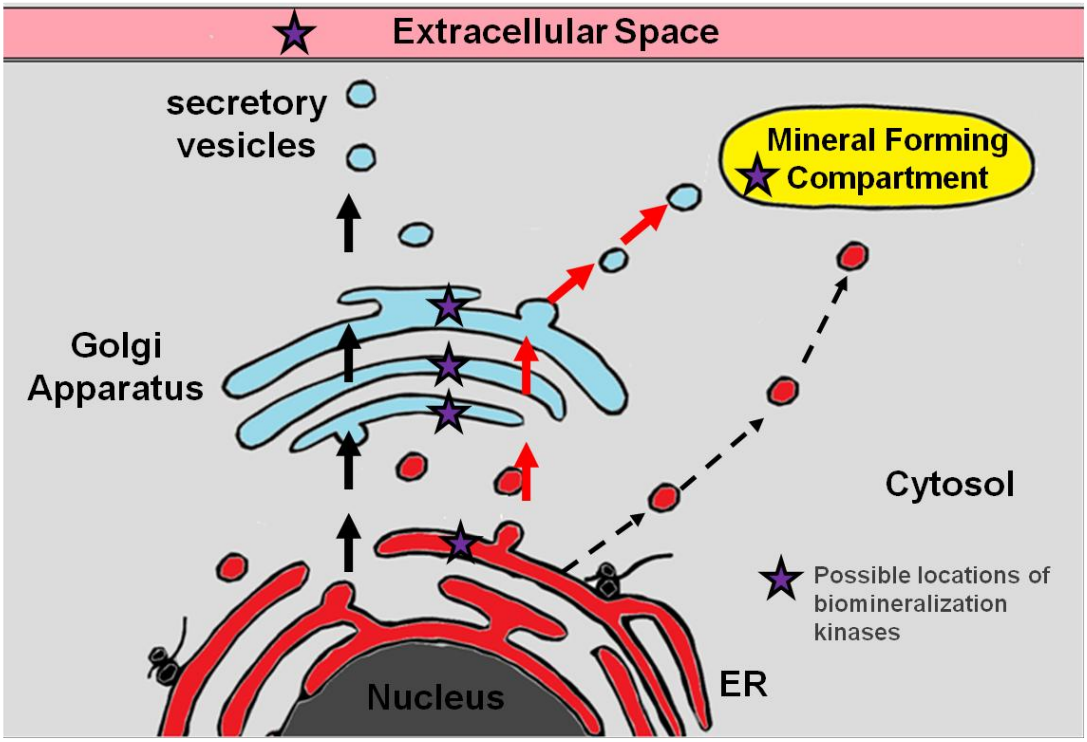


Figure 1.8. Model for extracellular sorting of proteins. Biomineral forming proteins are co-translationally imported into the endoplasmic reticulum (ER) via an N-terminal signal peptide. These proteins are then transported through the Golgi apparatus before reaching the biomineral forming compartments. Alternatively, biomineral forming proteins could travel from the ER and directly to the mineral forming compartment (dotted arrows).

Nearly all of the kinases characterized to date function within the cytosol with the exception of two mammalian kinases, Four-joint and Fyn. Four-joint (Fj) is a serine/threonine kinase from *Drosophila melanogaster* that phosphorylates extracellular cadherin domains of Fat protein and related proteins within the Golgi apparatus[83]. Fat is a type II transmembrane protein that, like other cadherins, requires calcium for cellular adhesion. The authors speculate that phosphorylation of Fat by Fj influences the interaction of the cadherin domains in Fat and which in turn affect the adhesion activity of the Fat protein[83]. Thus, Fj represents the first molecularly characterized kinase that functions in the secretory pathway. A second endomembrane kinase, Fyn, is a tyrosine kinase that phosphorylates Grp94 in the ER lumen of mammalian cells[84]. Phosphorylation of Grp94 by Fyn is required for the Grp94 mediated export of molecules involved in immune responses to the cell surface[84]. Although Fj and Fyn are kinases operating within the secretory pathway, neither Fj nor Fyn catalyzes the phosphorylation of proteins involved in biomineralization.

1.5 The Search for Diatom Biomineralization Kinases

As mentioned above, diatoms are suitable model organisms for studying biosilicification because: i) abundant material can be obtained from cultures grown to high density[1,12,51], ii) the complete genome sequence is available for some species[85,86], and iii) methods have been established for genetic manipulation[87,88]. Furthermore, identifying all of the components necessary for diatom silicification could eventually lead to the fabrication of biomimetic nanodevices (e.g. biosensors, microfluidic devices, drug

carriers) that exhibit similar intricately nanopatterned structures as the diatom biosilica at ambient conditions[48,50,89]. The diatom *Thalassiosira pseudonana* was the first diatom to have its genome sequenced making it a model diatom for studying diatom biosilica formation[87].

Despite previous research on the kinases involved in bone and teeth formation, there is a complete lack of knowledge of kinases involved in the phosphorylation of biomineralization proteins from non-mammalian systems such as diatoms. Given the importance of the phosphate moieties in diatom biomineral forming proteins, the kinases that phosphorylate these proteins may be critical components in controlling the biosilica structure produced. For example, it is possible that various hypo-phosphorylated forms of silaffins and silacidin possess different biomineral forming activities which in turn affect the biosilica structure (e.g. porosity, mechanical strength, elasticity).

Also of interest is the cellular localization of these biomineralization kinases which may provide some valuable insight into the intracellular transport of silaffins. The current hypothesis is that silaffins are co-translationally imported into the ER (via an N-terminal signal peptide) and then transported to the Golgi before reaching the silica deposition vesicle. This is supported by the presence of *O*-linked glycosylations which predominately occurs in the Golgi[60-61,90]. However, it cannot be excluded that silaffins are imported into the ER and then directly transported to the SDV by vesicles (see Figure 1.8). In this scenario, additional silaffin modifying enzymes (such as those that add *O*-linked glycosylations), may be present in the "silaffin transport" vesicles and not in the Golgi. Thus, if Golgi-resident kinases that phosphorylate silaffins and silacidin are discovered, then the current hypothesis of protein transport involved in biomineralization

would be substantiated.

Research dedicated to finding proteins involved in diatom biosilica formation has lead to the identification of several putative biomineralization kinases. Hildebrand and coworkers combined proteomics and transcriptomics to identify proteins potentially involved in diatom biomineralization[65]. After the addition of silicic acid to silicon-starved *Thalassiosira pseudonana* cells, proteins that were upregulated during silica formation (~4 to 6 hours after induction) were analyzed by mass spectroscopy and their corresponding genes were identified in the *T. pseudonana* genome database[65]. From this analysis, 31 proteins were found upregulated during biosilica formation, including two putative kinases, termed tpAK1 (gene id: newV2.0.grail.5.87.1) and tpSTK1 (gene id: newV2.0.grail.68.39.1). The mRNA expression of *tpak1* and *tpstk1* was also upregulated during silica biosynthesis and, in the case of *tpstk1*, the mRNA expression pattern was similar to that of silaffin *tpsil3*[65]. More recently, whole genome expression profiling of *T. pseudonana* under different nutrient limited conditions (i.e. silicon, iron, and nitrogen), at alkaline pH (pH = 9.4), and at low temperatures identified nearly 100 genes whose expression profile changed during silicon starvation[91]. Of these genes, the expression of *tpsil3* was downregulated along with nearly 10 genes encoding for putative kinases during silica starvation, suggesting a possible relationship of these kinases to the silaffin tpSil3.

Are the putative kinases found in the Hildebrand and Mock studies involved in the phosphorylation of silaffins and/or silacidins? Where are these kinases located within the *T. pseudonana* cell? The answer to these questions requires extensive biochemical and molecular cell biological analyses of the putative. This dissertation details the

structural and functional characterization of tpAK1 and tpSTK1 (the putative kinases from the Hildebrand study), and tpSTK2 (a putative tpSTK1-like kinase) to assess their involvement, if any, in silaffin and silacidin phosphorylation.

CHAPTER 2: EXPERIMENTAL PROCEDURE

2.1 Materials

Isopropyl-D-thiogalactopyranoside, guanidinium hydrochloride, sucrose, ATP, NADH, phenylmethylsulfonyl fluoride, and trichloroacetic acid were purchased from EMD Biosciences. Dephosphorylated casein, histones I and II, myelin basic protein, fluorenylmethyloxycarbonyl chloride, phosphoserine, phosphotyrosine, phosphothreonine, antimycin A, chymotrypsin, cytochrome c, inositol diphosphate, sodium deoxycholate, tetramethoxysilane, and Igepal and were purchased from Sigma-Aldrich. GTP and Tween-20 were purchased from Calbiochem. [γ-^{32}P]-ATP was purchased from MP Biomedicals. Sic1 (SDESEESEDSVSSEDEDWW) was synthesized by Genescript. Peptide R5 **(**SSKKSGSYSGSKGSKRRIL) was provided by Dr. Nils Kröger. Purified native silaffin tpSil3 and hydrogen fluoride treated native *Thalassiosira pseudonana* silaffins were provided by Dr. Nicole Poulsen.

2.2 Expression and isolation of recombinant silaffin substrates from *E. coli*[92]

Select regions of the silaffin genes *natsil1* from *Cylindrotheca fusiformis* and *tpsil1* and *tpsil3* from *T. pseudonana*, each containing a hexahistidine tag, were cloned into the pET16 vector (Novagen). For expression of recombinant silaffins, a freshly transformed *E. coli* DH5α clone harboring pET16b/rSilX (where X = C, N, 1L, or 3) was inoculated into LB medium containing 100 µg/mL ampicillin (LBamp), grown overnight at 37 ºC, diluted 50-fold with fresh LBamp medium and grown to an optical density (OD)

of 0.6 (at λ= 600 nm). Expression of the recombinant silaffins was induced by the addition of IPTG to a final concentration of 1 mM. After growing for an additional 3 hours at 37 ºC, the cells were harvested by centrifugation (4000*g*, 30 min, 4 ºC) and then washed once with 1 % (w/v) NaCl. Next, the cells were resuspended in *E. coli* lysis buffer (ECLB) containing 6 M Gdn-HCl (ECLB-G) and stirred for 15 minutes at 4 ºC. The cell mixture was sonicated (power setting 10, 30 sec on/off cycle) for 1 hour on ice and then centrifuged twice at 10,000*g* for 30 minutes at 4 ºC. Ni^{2+}-NTA resin (Qiagen) equilibrated in ECLB-G was added to the cleared lysate (supernatant), mixed by gentle rotation overnight at 4 ºC and then loaded onto a column. The Ni-NTA agarose was washed with 10 column volumes (CV) of ECLB containing 20 mM imidazole and the his-tagged protein was eluted stepwise with 4 CV of elution buffer (50 mM Tris-HCl pH 8.0, 1 M NaCl) containing increasing concentrations of imidazole (50 mM, 100 mM, 200 mM, and 1 M). All fractions were analyzed by SDS-PAGE[93], and fractions containing recombinant protein were pooled. Recombinant silaffin rSilC was subjected to an additional purification step using cation exchange chromatography. Briefly, Ni^{2+}-NTA fractions containing rSilC were dialyzed once against 50 mM ammonium acetate. High S cation exchange resin (BioRAD) equilibrated in 50 mM ammonium acetate was added to the dialysate, mixed by rotation for 2 hours at 4 ºC, and then loaded onto a column. After washing the resin with 10 CV of 50 mM ammonium acetate and then 10 CV of 2 M NH_3OH, rSilC was eluted with 5 CV of 2 M NaCl, 100 mM NH_3OH, 50 mM ammonium acetate. The eluate was immediately neutralized with the dropwise addition of acetic acid. All of the recombinant silaffins were dialyzed against 10 mM ammonium acetate for

three changes (~10 hr between changes) and then lyophilized. The dried protein residues were dissolved in milliQ H_2O.

2.3 Mass Spectroscopic Analysis of recombinant silaffins

The molecular mass of the recombinant silaffins were analyzed using a Micromass Quattro LC triple quadropole tandem mass analyzer. Samples were infused by a nanospray source in 50:50 water:acetonitrile containing 0.1 % formic acid. Mass spectroscopic analysis was performed at the Georgia Institute of Technology Bioanalytical Mass Spectrometry Facility.

2.4 Quantification of Protein Concentration

2.4.1 Bicinchoninic acid (BCA) Protein Assay[94]

Aliquots of protein solutions (from 0.5 - 20 μL) were diluted with water to a final volume of 25 μL in a 96-well plate. BSA samples ranging from 0 to 20 μg protein were also prepared to generate a standard curve. A 1:50 dilution of 4 % $CuSO_4$ in BCA reagent (Pierce) was prepared and 200 μL was added to each of the samples. The samples were incubated at 37 ºC for 30 minutes and then cooled at room temperature for 10 minutes. The absorbance was measured at 562 nm using a Synergy 2 Plate Reader (Biotek).

2.4.2 Amino Acid analysis[60]

2.4.2.1 Phenylisothiocyanate (PITC) derivatization

Amino acid standards were prepared as 100 mM stock solutions in H_2O. Lyophilized recombinant silaffins were subjected to acid hydrolysis for 24 hours at 110 ºC in 6 M HCl (Pierce) and then dried by lyophilization. The dried protein hydrolysates were dissolved in 50 µL ethanol/H_2O/Triethylamine (in a 2:2:1 ratio) and then dried under vacuum. The dried hydrolysate was dissolved in 50 µL ethanol/H_2O/Triethylamine/PITC (7:1:1:1) and then incubated at room temperature for 20 minutes. The PITC treated hydrolysate was dried under a vacuum and then dissolved in 50 µL 50 mM sodium phosphate buffer pH 6.4. Amino acid standards were derivatized with PITC as described above. After the final drying step, the PITC-derivatized amino acid standard was dissolved in 200 µL of 50 mM sodium phosphate buffer pH 6.4 to yield a final concentration of 0.5 nmol/µL.

2.4.2.2 Reverse phase-high pressure liquid chromatography (RP-HPLC)

The PITC-derivatized amino acid samples were separated using reverse phase HPLC on a Nucleosil C_{18} column (Machery-Nagel) by application of a linear acetonitrile gradient (buffer A: 50 mM sodium phosphate buffer pH 6.4, buffer B: 50 % acetonitrile, 50 mM sodium phosphate buffer pH 6.4; gradient 0-70 % Buffer B in 40 min). PITC-derivatized amino acids were detected at a wavelength of 254 nm. Relative amounts of PITC-derivatized amino acids from hydrolysates were determined by comparing the peak areas in the RP-HPLC chromatrogram to the peak areas from the standard amino acids.

2.5 Silica Precipitation Assays[58]

Recombinant silaffins were dissolved in 50 mM sodium phosphate citrate buffer (pH 4-10). A stock solution of monosilicic acid was prepared by hydrolyzing 1 M tetramethoxysilane (TMOS) in 1 mM HCl for 10 minutes at room temperature. Immediately after incubation, the monosilicic acid solution was added to the buffered silaffin solutions at a final concentration of 100 mM silicic acid, and then incubated at room temperature for 10 minutes. The precipitated silica was pelleted by centrifugation (16,000*g* for 5 min), washed three times with water, resuspended in 1 M NaOH, and then incubated at 95 ºC for 30 minutes. The concentration of silica in these solutions was quantified using the β-silicomolybdate method[95].

2.6 Culture Conditions

Thalassiosira pseudonana clone CCMP1335 was grown in an artificial seawater medium according to the North East Pacific Culture Collection[96] at 18 ºC under constant illumination at 10,000-15,000 lux.

2.7 Generation of a cDNA library coupled to oligo-dT_{25} magnetic beads

T. pseudonana cells ($5x10^7$ cells) were resuspended in 1 mL TRI-reagent (Sigma-Aldrich) and then gently lysed for 30 seconds using 0.1 mL of glass beads (0.25-0.30 mm, Sartorius). Total RNA was isolated from the lysed cells according to the

manufacturer's instructions. To remove contaminating DNA, the total RNA preparation was DNase treated in a final volume of 50 µL containing 1x One-Phor-All Plus buffer (GE Healthcare), 28.8 U RNA guard (GE Healthcare), and 1 U RNase-free DNase (Fermentas) for 30 minutes at 37 ºC. A second TRI-reagent RNA purification step was performed to remove the DNase and the total RNA concentration was measured spectrophotometrically at 260 nm and its quality was determined by agarose gel electrophoresis. Total RNA (5 µg) was coupled to oligo-dT_{25} magnetic Dynabeads (Invitrogen) and equilibrated in 2x binding buffer (20 mM Tris-HCl pH 8.0, 1.0 M LiCl, 2 mM EDTA all in diethylpyrocarbonate (DEPC) treated-H_2O) by incubating the RNA-bead mixture for 10 minutes at room temperature with constant shaking (1,400 rpm). The beads with attached *T. pseudonana* poly(A)+ RNA were washed with 10 mM Tris-HCl pH 8.0, 150 mM LiCl, 1 mM EDTA, and 0.05 % (v/v) Triton-X100 (Roche) three times and then incubated at 70 ºC in 1x First Strand Buffer (Invitrogen) for 10 minutes and then placed on ice. For cDNA synthesis, reverse transcription was performed in a final volume of 20 µL containing the oligo-dT_{25} beads with attached *T. pseudonana* poly(A)+ RNA, 0.5 mM dNTPs, 5 mM dithiothreitol (DTT), 1x First Strand Buffer, 28.8 U RNA guard, and 200 U Superscript III (Invitrogen) at 55 ºC for 50 minutes. After heat inactivation of the enzyme at 70 ºC for 15 minutes, mRNA was removed by the addition of 5 U RNaseH (Fermentas) and incubated at 37 ºC for 20 minutes. The beads with attached cDNA were washed twice with TE buffer containing 0.05 % (v/v) Triton-X100 and then resuspended in 20 µL TE buffer. This resulted in a *T. pseudonana* cDNA library coupled to the oligo $(dT)_{25}$ Dynabeads (*Tp* cDNA-beads).

2.8 Rapid amplification of cDNA ends (RACE) PCR: *tpak1* and *tpstk1*[97]

2.8.1 3' RACE

For amplification of the 3'-ends of the *tpak1* and *tpstk1* genes, two nested 3' RACE PCRs were performed using gene specific sense primers AK 5 (5'-GAA AGT TGC TAC CGA AGA GG-3') or STK 5 (5'-TGC AGC AGG GCA GTG CAT CC-3') for the first PCR and AK 6 (5'-GCC ATT TTG AAG GAG CAT GC-3') or STK 6 (5'-GCA TCC AAG CAG TAC ATT CC-3') for the second PCR. The first RACE PCR (94 °C 20 sec, 42 °C 20 sec, 72 °C 45 sec) was performed in a final volume of 50 µL containing 2 µL *Tp* cDNA-beads, 1x PCR Buffer, 0.1 mM dNTPs, 0.50 µM ON670 (5'-CAG CCG CCG AAT TCC CAG $(T)_{18}$-3'), 0.50 µM AK 5 or STK 5, and 0.25 µL Taq DNA polymerase (homemade). After 35 cycles of amplification, 1.0 µL of the reaction was used in a second PCR (35 cycles: 94 °C 20 sec, 55 °C 20 sec, 72 °C 30 sec) using sense primers AK 6 or STK 6 and antisense ON644 (5'-GCC GCC GAA TTC CCA GTT T-3'). The 550-bp and 500-bp DNA fragments produced from *tpak1* and *tpstk1* gene amplification, respectively, were cloned into the pGEMT vector (Promega) and sequenced.

2.8.2 5' RACE

The *T. pseudonana* cDNA-beads were prepared for 5' RACE by adding a poly C-tail to the 5'-end with terminal deoxynucleotidyltransferase (TdT) according to the manufacturer's instructions (Fermentas). The cDNA sequences of the 5'-end of the *tpak1* and *tpstk1* genes were amplified using two gene specific antisense primers: AK 14 (5'-

CAT AGT TCA ACG CCC TAC TCC-3') or STK 9 (5'-GTA GTG ATA CTG TCT CCA TC-3') for the first PCR and AK 13 (5'-GAT GAT TGT GGT GGA GTT GC-3') or STK 10 (5'-CCT CAC TGG TGA CGT AGT GA-3') for the second PCR. The first RACE PCR (94 °C 20 sec, 55 °C 20 sec, 72 °C 30 sec) was performed in a final volume of 50 µL containing 2 µL *Tp* cDNA-beads, 1x PCR Buffer, 0.1 mM dNTPs, 0.50 µM GIA anchor primer (5'-GGC CAC GCG TCG ACT AGT ACG GGI IGG GII GGG IIG-3'), 0.50 µM AK 14 or STK 9, and 0.25 µL Taq DNA polymerase. After 35 cycles of amplification, 1.0 µL of the reaction was used in a second PCR (35 cycles: 94 °C 20 sec, 55 °C 20 sec, 72 °C 30 sec) using the sense primers AK 13 or STK 10 and the antisense anchor primer R2 (5'-GGC CAC GCG TCG ACT AGT AC-3'). The 500-bp DNA fragments produced from *tpak1* and *tpstk1* gene amplification were cloned into the pGEMT vector (Promega) and sequenced.

2.9 Reverse Transcription (RT) PCR: tpAK1 and tpSTK1

2.9.1 tpAK1

The full length *tpak1* cDNA was amplified by PCR (40 cycles: 94 ºC 20 sec, 58 ºC 20 sec, 72 ºC 150 sec) using the sense primer (5'- ATG ACA GCT ACA CCG TAT AAC-3') and the antisense primer (5'-GTG AAG TAC AAT ACC CAA TGT CG-3') in a final volume of 50 µL containing 2.0 µL *Tp* cDNA-beads, 1x PCR Buffer, 0.1 mM dNTPs, 0.50 µM each primer, and 0.25 µL Taq DNA polymerase (homemade).

2.9.2 tpSTK1[97]

The full length *tpstk1* cDNA was amplified by PCR (40 cycles: 94 ºC 20 sec, 58 ºC 20 sec, 72 ºC 90 sec) using the sense primer 5'-ATG AGA GTC ATA CGC AAT GTT TC-3' and the antisense primer 5'-CTA AAA GCT GAT AGG ACC GAT G -3' in a final volume of 50 µL containing 2.0 µL *Tp* cDNA-beads, 1x PCR Buffer, 0.1 mM dNTPs, 0.50 µM each primer, and 0.25 µL Taq DNA polymerase (homemade).

2.10 Cloning, Expression, and Isolation of recombinant tpAK1

2.10.1 Cloning of recombinant tpAK1

The coding region of *tpak1* was amplified from *Tp* cDNA-beads by PCR (25 cycles: 94 °C 20 sec, 58 °C 20 sec, 72 °C 150 sec) using the sense primer AK 5'-**CAT ATG** ACA GCT ACA CCG TAT AAC AGG-3' and the antisense primer 5'-*GGA TCC* CTA ATG GTG ATG GTG ATG GTG ACC TTC AAG TTG GGC CT-3', which introduced an *Nde*I restriction site (bold), a *Bam*HI site (italic), and C-terminal hexahistidine tag (underlined), in a final volume of 50 µL containing 2.0 µL *Tp* cDNA-beads, 1x PCR Extender buffer, 60 µM dNTPs, 0.50 µM each primer, and 2.5 U PCR Extender DNA polymerase mix (5 PRIME). The resulting 2.3-kb PCR was ligated into the pGEMT vector (Promega) and sequenced. Plasmid pGEMT/tpAK1 was digested with *Nde*I and *Bam*HI and the resulting 2.3-kb DNA fragment was ligated into the *Nde*I/*Bam*HI sites of the pET11a vector (Novagen) and introduced into *E. coli* DH5α.

2.10.2 Expression and Isolation of recombinant tpAK1

Recombinant tpAK1, from an *E. coli* BL21 (DE3) clone harboring pET11a/tpAK1, was expressed as described for the recombinant silaffins (section 2.2). After the NaCl wash, the cells were resuspended in ECLB at 5 mL buffer/gram wet pellet weight. Lysozyme was added to a final concentration of 1 mg/mL and the mixture was incubated on ice for 30 minutes. The cells were lysed by sonication (power setting 10, 30 sec on/off cycle for 1 hr) using a Misonix Sonicator 3000 and then centrifuged at 12,000*g* for 30 minutes at 4 ºC. The cell pellet was washed with 10 mL of ECLB containing 3 M urea and the mixture was sonicated as described above. After centrifugation of the lysate at 12,000*g* for 30 minutes at 4 ºC, the pellet was resuspended in ECLB containing 8 M urea (ECLB-U) and mixed by gentle rotation at 4 ºC for 30 minutes to solubilize inclusion bodies. The mixture was centrifuged at 12,000*g* for 30 minutes at 4 ºC. Ni^{2+}-NTA resin (Qiagen) equilibrated in ECLB-U was added to the solubilized inclusion bodies and mixed by rotation overnight at 4 ºC and then loaded onto a column. The resin was washed with 10 column volumes (CV) of ECLB containing 20 mM imidazole and 8 M urea. Recombinant tpAK1 was eluted stepwise with 4 CV of ECLB-U containing increasing concentrations of imidazole (50 mM, 100 mM, 200 mM, and 1 M). All of the fractions were analyzed by SDS-PAGE[93] and fractions containing recombinant tpAK1 were combined. Recombinant tpAK1 was refolded by dialysis against tpAK1 refolding buffer (50 mM Tris-HCl pH 7.0, 0.1 M NaCl, 10 % (v/v) glycerol, and 1 mM PMSF) containing decreasing concentrations of DTT (1 mM, 0.5 mM, 0.2 mM, 0 mM), EDTA (0.5 M, 0.25 M, 0.1 M, 0 M) and urea (8 M, 4 M, 2 M, 0 M), respectively, at 4 ºC for 10

hours in each buffer. The dialysate was concentrated to ~4 mg/mL of recombinant tpAK1 using a 10,000 MWCO centricon filter (Millipore).

2.11 Semiquantitative RT-PCR from Synchronized *T. pseudonana* Cells[97-98]

Sodium silicate (420 μM) was added to a 1.0-1.2 x 10^6 cells/mL *T. pseudonana* cell culture to achieve a high cell density (3.5-4.0 x 10^6 cells/mL). The cells were pelleted by centrifugation (3,000*g* for 10 min), washed once in silicate free-artificial seawater medium (Si-ASW), and inoculated into Si-ASW medium in a polycarbonate bottle to a final density of 1.0 x 10^6 cells/mL. The culture was stirred and aerated for 24 hours. Cell cycle progression was initiated by the addition of sodium silicate to a final concentration of 200 μM. At the indicated times equal culture volumes were harvested for total RNA isolation and cDNA synthesis (see section 2.7) and for Western Blot analysis (see section 2.20). The *tpstk1* gene was amplified using sense primer 5'-AAG CTT TCA GCT CAA CTT ATA CTT GCT CCC-3' and antisense primer 5'-GAG CAG GCC TTC ACA TCA CAA CGC ACC GTC-3'. The *tpsil3* gene was amplified using sense primer 5'-GGA CAC CAA GAG TGG AAA G-3' and antisense primer 5'-TCA AGC GCT CAT GGA GTG G-3'. The *tpsil1* gene was amplified using sense primer 5'-ATG AAA GTT ACC ACG TCA ATC-3' and antisense primer 5'-AGA TCG CGA CGC AAC ATT C-3'. The *tpsic* gene was amplified using sense primer 5'-CGG ACT TTG TTG ATA GAT GAA CC-3' and antisense primer 5'-ATG GTC AAG TAC AAC GTC CTC G-3'. The *tpfcp9* gene was amplified using sense primer 5'-TTC GAG GAT GAG CTC GGT G-3' and antisense primer 5'-GAA CAT CAC GCA TGA AGG C-3'. PCRs were performed with

25 cycles of amplification (94 °C 20 sec, 60 °C 20 sec, and 72 °C 40 sec) in a final volume of 50 µL containing *Tp* cDNA-beads (0.25 -1 µL), 1x PCR Buffer, 0.1 mM dNTPs, 0.50 µM each primer, and 0.25 µL Taq DNA polymerase (homemade).

2.12 Genomic DNA isolation from *T. pseudonana*

T. pseudonana wild type cells (2 x 10^8 cells) were harvested by centrifugation (3,000*g*, 10 min, 18 ºC) and resuspended in 600 µL Nuclei lysis solution (Promega). Glass beads (100 µL) were added and the mixture was vortexed for 30 seconds. The mixture was heated at 65 ºC for 15 minutes and then cooled to room temperature before adding 1.5 µL of 10 mg/mL RNase A. Following a 15 minute incubation at 37 ºC, 250 µL of Protein precipitation solution (Promega) was added to the mixture. The mixture was centrifuged at 16,000*g* for 30 minutes at 4 ºC and the supernatant was transferred to a fresh 1.5 mL eppendorf tube. Isopropanol (700 µL) was added to the supernatant and the mixture was incubated on ice for 20 minutes. The sample was centrifuged at 16,000*g* for 1 minute at 4 ºC and the pellet was washed with 600 µL 70 % (v/v) ethanol. After centrifugation (16,000*g*, 1 min, 4 ºC), the pellet was air-dried for 15 minutes, resuspended in 50 µL of TE buffer and incubated at 65 ºC for 15 minutes. The *T. pseudonana* genomic DNA (*Tp* gDNA) concentration was estimated spectrophotometrically at 260 nm and its purity determined by agarose gel electrophoresis.

2.13 Cloning, Expression, and Isolation of recombinant tpSTK1[97]

2.13.1 Cloning of recombinant tpSTK1

As the *tpstk1* gene was shown to contain no introns, the coding region of *tpstk1* was amplified from *Tp* gDNA by PCR (20 cycles: 94 °C 20 sec, 58 °C 20 sec, 72 °C 90 sec) using the sense primer 5'-GC**A GGC CT**A TGA GAG TCA TAC GCA ATG TTT CTC-3' and the antisense primer 5'-CC*A AGC TT*C TAA TGG TGA TGG TGA TGG TGA AAG CTG ATA GGA CCG ATG CC-3', which introduced a *Stu*I restriction site (bold), a *Hind*III site (italic), and C-terminal hexahistidine tag (underlined), in a final volume of 50 µL containing 400 ng *Tp* gDNA, 1x PCR Extender buffer, 60 µM dNTPs, 0.50 µM each primer, and 2.5 U PCR Extender polymerase mix (5 PRIME). The resulting 1.5-kb PCR product was digested with *Stu*I and *Hind*III, ligated into the *Stu*I/*Hind*III sites of the pPROEX-HTb vector (Life Technologies) and introduced into *E. coli* DH5α. By this strategy a gene was generated that encodes the full length tpSTK1 protein fused to an N-terminal hexahistidine tag and followed by a spacer region (DYDIPTT), an rTEV protease cleavage site, and a C-terminal hexahistidine tag. This fusion protein was termed recombinant tpSTK1.

2.13.2 Expression and Isolation of recombinant tpSTK1

Recombinant tpSTK1, from an *E. coli* BL21 (DE3) clone harboring pPROEX-HTb/tpSTK1, was expressed as described for the recombinant silaffins (section 2.2). After the NaCl wash, the cells were resuspended in lysis buffer A (50 mM Tris-HCl pH

8.0, 1 mM EDTA, 10 mM DTT, 25 % (w/v) sucrose). Lysozyme, DNase I, and $MgCl_2$ were added at final concentrations of 0.5 mg/mL, 19 µg/mL, and 2.0 mM, respectively. Subsequently, an equal volume of lysis buffer B (50 mM Tris-HCl pH 8.0, 0.1 M NaCl, 10 mM DTT) was added and the cells were incubated at room temperature for 1 hour. EDTA was added to a final concentration of 6.7 µM and the cells were frozen in liquid nitrogen. The cell pellet was thawed for 30 minutes at 37 ºC. $MgCl_2$ was added to a final concentration of 4.1 mM and the cells were incubated at room temperature for 1 hour. EDTA (6.7 µM) was added and the lysate was centrifuged at 23,000*g* for 30 minutes at 4 ºC to pellet inclusion bodies. The pellet was resuspended in 10 mL of wash buffer I (50 mM Tris-HCl pH 8.0, 1 mM EDTA, 10 mM DTT, 0.1 M NaCl) and then centrifuged at 23,000*g* for 30 minutes at 4 ºC. The pellet was resuspended with 10 mL of wash buffer II (50 mM Tris-HCl pH 8.0, 1 mM EDTA, 0.1 M NaCl, 1 mM DTT) and centrifuged at 23,000*g* for 30 minutes at 4 ºC. The final pellet was dissolved in purification buffer (50 mM Tris-HCl pH 8.0, 1 mM DTT, 0.5 M NaCl, 20 mM imidazole, 8 M urea) for 30 minutes, and applied to a 1 mL HisTrap-FF column (GE Healthcare) (pre-equilibrated with purification buffer) at 4 ºC for 1 hour at 0.5 mL/min. The column was washed with 25 mL of purification buffer before refolding of the protein was performed on the column with a linear gradient (0.5 mL/min, 50 min) from 100 % purification buffer to 100 % tpSTK1-refolding buffer (50 mM Tris-HCl pH 8.0, 0.5 M NaCl, 20 mM imidazole) at 4 ºC. After 25 mL washing with refolding buffer the protein was eluted with a linear gradient (0.5 mL /min, 20 min) from 100 % refolding buffer to 100 % elution buffer (50 mM Tris-HCl pH 8.0, 0.5 M NaCl, 1 M imidazole). Fractions containing recombinant tpSTK1, as determined by SDS-PAGE[93] analysis, were dialyzed against 50 mM Tris-HCl

pH 8.0, 0.1 M NaCl, 10 % (v/v) glycerol, and 1 mM PMSF. The dialysate was concentrated to ~2 mg/mL recombinant tpSTK1 using a 3,000 molecular weight cut-off (MWCO) centricon filter (Millipore).

2.14 Preparation of anti-tpSTK1 IgG[97]

Recombinant tpSTK1 (5 mg) was excised from a preparative SDS-PAG (Gel size: 13 x 13 cm^2, 2 mm thickness). The excised gel slice containing the recombinant protein was cut into several pieces. The pieces were crushed by passing them through a 1.0 mL syringe. Recombinant tpSTK1 was passively eluted from the crushed gel pieces by incubation in water overnight at 4 ºC with constant stirring. The mixture was then centrifuged at 100,000*g* for 30 minutes to pellet gel fragments. The protein solution (supernatant) containing recombinant tpSTK1 was dried by lyophilization, the protein residue reconstituted in complete Freud's adjuvant (source: *Mycobacterium tuberculosis*) and then injected into rabbits for polyclonal antibody production (Lampire Biological Laboratories). A tpSTK1 affinity matrix was generated by coupling 3.5 mg of SDS-PAG purified recombinant tpSTK1 to N-hydroxysuccinimide-activated HiTrap sepharose according to the manufacturer's instructions (GE Healthcare). The rabbit antisera obtained 50 days after immunization was diluted 1:1 with PBS and applied to the tpSTK1 affinity column via a peristaltic pump (setting 4, speed 1x) and re-circulated overnight at 4 ºC. The column was washed with PBS until the absorbance at 280 nm was less than 0.05. The antibodies were eluted with 100 mM glycine pH 2.5 and fractions were collected dropwise into 5 µL of 1 M Tris-HCl pH 9.0 (required for immediate pH

neutralization). IgGs from affinity matrix purified anti-tpSTK1 and from the preimmune sera were obtained using a Protein G column (GE Healthcare) according to the manufacturer's instructions.

2.15 Western Blot analysis

Samples were subjected to SDS-PAGE[93] and then the gel was placed onto a nitrocellulose membrane presoaked in Tris-glycine transfer buffer pH 8.3 (25 mM Tris, 192 mM glycine, 20 % (v/v) methanol, 0.025 % (w/v) SDS). Electrotransfer was performed using a wet transfer apparatus (BioRAD) at 100 volts for 1 hour. After transfer, the membrane was washed twice for 10 minutes in TBS containing 0.05 % (v/v) Tween-20 (TBST) and then incubated in TBS SuperBlock buffer (Pierce) for 30 minutes with rocking. The membrane was washed three times for 10 minutes in TBST buffer and then incubated with 0.2 µg/mL anti-tpSTK1 (or 0.3 µg/mL anti-tpSTK2; see section 2.21.2) polyclonal antibody in blocking buffer for 1 hour at room temperature. The membrane was washed twice with TBST and then incubated with anti-rabbit IgG horseradish peroxidase conjugate (1:10,000 dilution in TBST, Sigma-Aldrich) for 1 hour at room temperature with gentle rotation. After the incubation, the membrane was washed four times with TBST buffer and then incubated with SuperSignal Working Solution (Pierce) according to the manufacturer's instructions. After incubation, the blot was placed in clear, colorless plastic wrap and exposed to X-ray film (Kodak). When using anti-GFP as the primary antibody (0.5 µg/mL, ClonTech), PBS containing 0.05 % (v/v)

Tween-20 (PBST) was used instead of TBST and the blocking buffer was 2 % (w/v) bovine serum albumin (BSA) in PBST.

2.16 Kinase Assays

2.16.1 Radioactive assays

Protein substrates (20 µg) were incubated with 10 µg of recombinant tpSTK1 or recombinant tpAK1 in kinase activity buffer (50 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$, and 10 µM ATP) with 2 µCi [γ-^{32}P]-ATP for 1 hour at 37 ºC. To test for divalent metal ion requirement, $MgCl_2$ was replaced by equal concentrations of $CaCl_2$, $MnCl_2$, or $ZnCl_2$. Reactions were stopped by the addition of an equal volume of SDS sample buffer, incubated at 95 ºC for 15 minutes and run on a 15 % SDS-PAG. Following SDS-PAGE, the gel was fixed in 30 % (v/v) methanol, 7 % (v/v) acetic acid for 10 minutes at room temperature, washed with water, dried, and then exposed to an imaging plate (Fujifilm) overnight. The imaging plate was analyzed with a Fuji Fluorescence/Phosphor-imager.

2.16.2 Luminescence based assays[97,99]

Protein and peptide substrates (10 µg) were incubated with 10 µg of recombinant tpSTK1 (or tpSTK2-KD) in kinase activity buffer (50 mM Tris-HCl pH 8.0, 10 mM $MgCl_2$, and 10 µM ATP) for 1 hour at 37 ºC. For recombinant tpSTK1, the test for divalent metal ion requirement was performed as described in section 2.16.1. For detection an equal volume of Luminescent Kinase Assay reagent (Kinase-Glo® kit,

Promega) was added to the reaction and incubated at room temperature for 10 minutes. Luminescence was measured using a Synergy 2 Plate Reader (Biotek).

2.17 Recombinant silacidin constructs

2.17.1 Construction of pET28a/rSic

A DNA molecule corresponding to nucleotides 154-730 of the *silacidin* (*tpsic*) gene[62] was amplified from *Tp g*DNA by PCR (20 cycles: 94 °C 20 sec, 58 °C 20 sec, 68 °C 60 sec) using the sense primer 5'-CAT G**CC ATG G**GA ATC AAC AAA GTC CGC CGT CTC-3' and the antisense primer 5'-GCC G*CT CGA G*CT AGT GAT GGT GAT GGT GAT GAT CAA ACA TCA TCA AAT CTT CAC T-3', which introduced an *Nco*I site (bold), an *Xho*I site (italic), and a C-terminal hexahistidine tag (underline), in a final volume of 50 µL containing 500 ng *Tp* gDNA, 1x Pfx buffer, 60 µM dNTPs, 0.50 µM each primer, 1.0 mM $MgSO_4$, and 1.0 U Pfx DNA polymerase (Invitrogen). The resulting 620-bp product, which encodes amino acids 52-184 of the silacidin precursor polypeptide, was digested with *Nco*I and *Xho*I, ligated into the *Nco*I and *Xho*I sites of pET28a and introduced into *E. coli* DH5α yielding expression plasmid pET28a/rSic. The DNA sequence of r*Sic* in the pET vector was verified by sequencing.

2.17.2 Construction of $rSilN_{H10}$-rSic fusion[97]

2.17.2.1 Cloning of $rSilN_{H10}$

rsilN was amplified from pET16b/rSilN by PCR (20 cycles: 94 °C 20 sec, 60 °C 20 sec, 68 °C 45 sec) using the sense primer 5'-CAT G**CC ATG G**CT GCC CAA AGC

ATT GCT GAC-3' and the antisense primer 5'-CGC G*GG ATC C*GT GAT GGT GAT GGT GAT GGT GAT GGT GAT GCA AGA TAC GAA GTT CTT CCT C-3', which introduced an *Nco*I restriction site (bold), a *BamHI* site (italic), and C-terminal decahistidine tag (underlined), in a final volume of 50 µL containing 500 ng pET16b/rSilN, 1x Pfx buffer, 60 µM dNTPs, 0.50 µM each primer, 1.0 mM $MgSO_4$, and 1.0 U Pfx DNA polymerase (Invitrogen). The 310-bp PCR product obtained was digested with *Nco*I and *Bam*HI, ligated into the *Nco*I/*Bam*HI sites of the pET28a vector (Novagen) and introduced into *E. coli* DH5α, yielding plasmid pET28a/rSilN_{H10}.

2.17.2.2 Cloning of rSilN_{H10}-rSic

The *tpsic* gene was amplified as described (section 2.10.1) except the sense primer 5'-CGC G**GG ATC C**AT GAT CAA CAA AGT CCG CCG TCT C-3'and the antisense primer 5'-GCC G*CT CGA G*CT ATG CAT CAA ACA TCA TCA AAT CTT CAC T-3', which introduced a *Bam*HI site (bold) and an *Xho*I site (italic), was used in the PCR reaction. The resulting 596-bp product was ligated into pJET1.2 (Fermentas) yielding plasmid pJET1.2/rSic. After confirming the *rsic* gene sequence by DNA sequencing, plasmid pJET1.2/rSic was digested with *Bam*HI and *Xho*I, and the 596-bp product was ligated into the *Bam*HI and *Xho*I sites of pET28a/rSilN_{H10} and then introduced into *E. coli* DH5α yielding expression plasmid pET28a/rSilN_{H10}-rSic. The composition and correct orientation of the rSilN_{H10}-rSic fusion gene was verified by restriction site mapping.

2.17.3 Expression and isolation of recombinant silacidin constructs

Recombinant silacidin constructs (rSic and rSilN_{H10}-rSic) were expressed as described for the recombinant silaffins (section 2.2) except cells were grown in medium containing 100 μg/mL kanamycin instead of LB-amp. The silaffin-silacidin fusion protein (rSilN_{H10}-rSic) was purified via Ni^{2+}-NTA chromatography as described for the recombinant silaffins.

2.18 Cyanogen bromide cleavage of rSilN_{H10}-rSic[97]

Cyanogen Bromide (Acros Organics) dissolved in 70 % (v/v) formic acid was added to lyophilized rSilN_{H10}-rSic at 1 mg CNBr per 0.1 mg dry protein. The reaction mixture was incubated at room temperature in the dark for 24 hours. Excess CNBr was removed by a steady stream of N_2 gas, and the residue was dissolved in 0.2 mL H_2O and lyophilized. The dry sample was resuspended in 1 M Tris, neutralized with 2 M NaOH, and incubated with Ni^{2+}-NTA sepharose (GE Healthcare) in 50 mM Tris-HCl pH 8.0, 1 M NaCl, 10 mM imidazole. The unbound material contained pure rSic, which was exhaustively dialyzed against 10 mM ammonium acetate. The dialysate was dried by lyophilization and the residue dissolved in H_2O.

2.19 Reverse phase high performance liquid chromatography (HPLC) of silaffins[97]

Recombinant silaffins before and after incubation with recombinant tpSTK1 were separated using reverse-phase HPLC on a Nucleosil C_{18} column (Machery-Nagel) by

application of a linear acetonitrile gradient (buffer A: 0.1 % (v/v) trifluoroacetic acid (TFA) in H_2O, buffer B: 0.085 % (v/v) TFA in acetonitrile; gradient 0-25 % Buffer B in 35 min). Proteins were detected at a wavelength of 226 nm.

2.20 Chemical analysis of phosphorylated silaffins

2.20.1 Phosphate release determination (malachite green method)[100]

Lyophilized protein samples were resuspended in 25 µL of 10 % (w/v) $MgNO_3 \bullet 6H_2O$ (in 95 % (v/v) ethanol) and incubated at 95 ºC for 20 minutes and then ashed. After cooling to room temperature (RT), 200 µL of 1.2 M HCl was added to the sample, vortexed, and then incubated at RT for 20 minutes. Aliquots of the sample were added to 600 µL 1.2 M HCl and 200 µL of the phosphate reagent (3 parts 0.2 % (w/v) malachite green with 1 part 10 % (w/v) $(NH_4)_6Mo_7O_{24}.4H_2O$), incubated at RT for 10 minutes, and the absorbance at 660 nm was measured with a Shimadzu spectrophotometer. A standard curve was generated using 0 to 6 nmol KH_2PO_4.

2.20.2 Phosphoamino acid analysis[101]

2.20.2.1 Purification of phosphoamino acids

Phosphoserine, phosphotyrosine, and phosphothreonine were prepared as 1 mM stock solutions in H_2O. Lyophilized native tpSil3 and tpSTK1 phosphorylated rSil3 (100 nmol based on phosphate content) were subjected to partial acid hydrolysis for 3 hours at 110 ºC in 6 M HCl (Pierce) and then dried by lyophilization. The dried protein hydrolysates were dissolved in 0.3 mL of 100 mM formic acid and passed through a

column containing 0.35 mL Dowex 50W-X8 (H+) cation exchange resin (BioRAD) equilibrated in 100 mM formic acid. The resin was then washed twice with 0.8 mL of 100 mM formic acid. The flow-through and washes containing the phosphoamino acids were lyophilized.

2.20.2.2 Fluorenylmethyloxycarbonyl (FMOC) derivatization

The enriched phosphoamino acids were resuspended in 250 µL of 0.2 M H_3BO_3-NaOH pH 6.2. The phosphoamino acids were derivatized by the addition of 250 µL of 15 mM fluorenylmethyloxycarbonyl-HCl (FMOC-Cl), vortexed briefly, and the reaction was allowed to proceed for 60 seconds. To remove unreacted FMOC-Cl, the mixture was extracted with *n*-pentane. After allowing the phases to separate, the organic phase was removed by aspiration and the pentane extraction was repeated twice. In the last extraction step, the aqueous layer (bottom layer) was removed using a 1 mL syringe equipped with a pointed end 27.75 gauge needle. Phosphoamino acid standards (100 nmol) were derivatized with FMOC as described above.

2.20.2.3 Anion Exchange HPLC

The FMOC-derivatized phosphoamino acid samples were subjected to high pressure anion exchange chromatography at a flow rate of 1.5 mL/min on a Partisil 10 SAX WCS analytical column (Whatman). Separation of the FMOC-derivatized amino acids was performed with an isocratic elution in pH 3.9 buffer containing 55 % (v/v) methanol, 1 % (v/v) tetrahydrofuran, and 10 mM potassium phosphate at room temperature on the Agilent 1100 Series HPLC system. FMOC-derivatized phosphoamino acids were detected with a UV detector at 265 nm.

2.21 Cloning, Expression, and Isolation of recombinant tpSTK2 Kinase Domain

2.21.1 Cloning of recombinant tpSTK2-KD

The coding region of the predicted kinase domain of *tpstk2* (*tpstk2-KD*) was amplified from *Tp* cDNA-beads by PCR (20 cycles: 94 °C 20 sec, 60 °C 20 sec, 68 °C 60 sec) using the sense primer 5'- CAT G**CC ATG G**AC CTG ACT GGT AAA ACT ATC-3' and the antisense primer 5'-GGA ATT C*CA TAT G*TT AAT GGT GAT GGT GAT GGT GGA TGG ATG ATC CAA ATA AGA AG-3' which introduced an *Nco*I restriction site (bold), an *Nde*I site (italic), and C-terminal hexahistidine tag (underlined), in a final volume of 50 µL containing 2 µL *Tp* cDNA-beads, 1x Pfx buffer, 60 µM dNTPs, 0.50 µM each primer, 1.0 mM $MgSO_4$, and 1.0 U Pfx DNA polymerase (Invitrogen). The resulting 0.93-kb PCR product was digested with *Nco*I and *Nde*I, ligated into the *Nco*I/*Nde*I sites of the pET16b vector (Novagen) and introduced into *E. coli* DH5α yielding plasmid pET16b/tpSTK2-KD. The DNA sequence of *tpstk2-KD* in the pET vector was verified by sequencing.

2.21.2 Expression and isolation of recombinant tpSTK2-KD

Recombinant tpSTK2 kinase domain (from an *E. coli* BL21 (DE3) clone harboring pET16b/tpSTK2-KD) was expressed, purified from inclusion bodies, and refolded as described for recombinant tpSTK1 (section 2.13.2). Recombinant tpSTK2-KD (6 mg) was purified from a preparative SDS-PAG and a tpSTK2 affinity column was prepared with 3.5 mg of pure recombinant tpSTK2-KD to purify anti-tpSTK2 IgGs from

rabbit antisera obtained 60 days after immunization (against recombinant tpSTK2-KD) (see section 2.14 for details).

2.22 Expression of GFP fusion proteins

2.22.1 GFP_{DDEL}[97]

To generate a gene encoding cytosolically expressed GFP with a C-terminal ER retention signal (GFP_{DDEL}), the *egfp* gene was amplified with the sense primer 5'-ACC AAA AAT GGT GAG CAA GGG CGA GGA C-3' and antisense primer 5'-GAT TC**G CGG CCG C**TT ACA GCT CGT CAT CCT TGT ACA GCT CGT CCA TGC C-3', which introduced a *Not*I restriction site (bold) and the tetrapeptide DDEL (underlined) in a final volume of 50 µL containing 500 ng pTpfcp/ctGFP[87], 1x Pfx buffer, 60 µM dNTPs, 0.50 µM each primer, 1.0 mM $MgSO_4$, and 1.0 U Pfx DNA polymerase (Invitrogen). After 20 cycles of amplification (94 °C 20 sec, 60 °C 20 sec, 68 °C 45 sec) the resulting 750-bp PCR product was digested with *Not*I and ligated into the *Eco*RV and *Not*I sites of the vector pTpfcp generating pTpfcp/GFP_{DDEL}.

2.22.2 $tpSTK_{1-24}GFP_{DDEL}$[97]

To generate a fusion gene encoding the *tpstk1* putative signal peptide (amino acids 1-24) fused to a C-terminal GFP tag, a region of the *tpstk1* gene (1-72) was amplified by PCR (20 cycles: 94 °C 20 sec, 60 °C 20 sec, 68 °C 20 sec) using the sense primer 5'-ACC AAA ATG AGA GTC ATA CGC AAT G-3' and the antisense primer 5'-ATT CGG TAC CGG AGG TGA CGG ACA CAA GAA A-3' in a final volume of 50

µL containing 200 ng *Tp* gDNA, 1x Pfx buffer, 60 µM dNTPs, 0.50 µM each primer, 1.0 mM $MgSO_4$, and 1.0 U Pfx DNA polymerase (Invitrogen). The resulting 88-bp PCR product was ligated into the *Eco*RV site of pTpfcp/ctGFP[87], generating pTpfcp/tpSTK1$_{1-24}$GFP. The orientation of the *tpstk1* DNA fragment in the vector was verified by PCR (35 cycles: 94 °C 20 sec, 55 °C 20 sec, 72 °C 30 sec) using a *tpstk1* gene specific sense primer 5'- ACC AAA ATG AGA GTC ATA CGC AAT G -3') and the GFP 37 gene specific antisense primer in a final volume of 50 µL containing 500 ng pTpfcp/tpSTK1$_{1-24}$GFP, 1x Dream Taq Buffer, 0.1 mM dNTPs, 0.50 µM each primer, and 1 U Dream Taq DNA polymerase (Fermentas). To introduce a C-terminal ER-retention signal on the pTpfcp/STK1$_{1-24}$GFP gene fusion, plasmid pTpfcp/tpSTK1$_{1-24}$GFP was then used as a template for PCR amplification of a tpSTK1$_{1-24}$GFP$_{DDEL}$ encoding gene using the sense primer 5'-ACC AAA ATG AGA GTC ATA CGC AAT G-3' and the antisense primer 5'-GAT TC**G CGG CCG C**TT A<u>CA GCT CGT CAT C</u>-3', which introduced a *Not*I restriction site (bold) and the diatom ER retention tetrapeptide DDEL (underlined) in a final volume of 50 µL containing 500 ng plasmid, 1x Pfx buffer, 60 µM dNTPs, 0.50 µM each primer, 1.0 mM $MgSO_4$, and 1.0 U Pfx DNA polymerase (Invitrogen). The resulting 810-bp product was digested with *Not*I and ligated into the *Eco*RV and *Not*I site of the previously described vector pTpfcp generating pTpfcp/tpSTK1$_{1-24}$GFP$_{DDEL}$.

2.22.3 tpSTK1-GFP

The *tpstk1* gene was amplified from *T. pseudonana* genomic DNA by PCR using the sense primer 5'-**ACC AAA** ATG AGA GTC ATA CGC AAT G-3', which introduced 6-bp from the promoter region upstream of the start codon (bold), and the antisense

primer 5'-AAA GCT GAT AGG ACC GAT GCC-3' in a final volume of 50 µL containing 200 ng *Tp* gDNA, 1x Pfx buffer, 60 µM dNTPs, 0.50 µM each primer, 1.0 mM $MgSO_4$, and 1.0 U Pfx DNA polymerase (Invitrogen). The 1.5-kb product obtained after 20 cycles of PCR amplification (94 °C 20 sec, 60 °C 20 sec, 68 °C 90 sec) was blunt-end ligated into the *Eco*RV site of plasmid pTpfcp/ctGFP[87] to generate plasmid pTpfcp/tpSTK1GFP. The orientation of *tpstk1* in the vector was confirmed by restriction mapping and the sequence confirmed by DNA sequencing.

2.22.4 BiPGFP$_{DDEL}$

The *tpbip* gene was amplified from *Tp* gDNA by PCR using the sense primer 5'-ATC ATA ATC ATG GCG TTC AAA CGG CGG TTC-3'and the antisense primer 5'-GGT GGC GGA GAC GAT AAT GTC GGA TTT-3' was in a final volume of 50 µL containing 200 ng *Tp* gDNA, 1x Pfx buffer, 60 µM dNTPs, 0.50 µM each primer, 1.0 mM $MgSO_4$, and 1.0 U Pfx DNA polymerase (Invitrogen). The 2.03-kb product obtained after 20 cycles of PCR (20 cycles: 94 °C 20 sec, 60 °C 20 sec, 72 °C 130 sec) amplification was blunt-end ligated into the *Eco*RV site of plasmid pTpfcp/ctGFP$_{DDEL}$ to generate plasmid pTpfcp/tpBiP$_{DDEL}$. The orientation of *tpbip* in the vector was verified by restriction mapping and the sequence confirmed by DNA sequencing.

2.22.5 GFP-tpSTK1

The *tpstk1* gene was amplified from *Tp* gDNA as described in section 2.22.3, except the antisense primer 5'- CCA AGC TTC TAA AAG CTG ATA GGA CCG ATG CC -3' was used in the PCR reaction. The 1.5-kb product obtained after 20 cycles of PCR

amplification was blunt-end ligated into the *Eco*RV site of plasmid pTpfcp, in which the *Kpn*I site had been destroyed, to generate plasmid pTpfcp(-*Kpn*I)/tpSTK1. The *tpstk1* sequence was confirmed by DNA sequencing. A second PCR (20 cycles: 94 °C 20 sec, 59 °C 20 sec, 68 °C 45 sec) was performed on an eGFP-containing plasmid with the sense primer GFP 40 (5'-GC**G GTA CC**A TGG TGA GCA AGG GCG AGG A-3') and the antisense primer GFP 41 (5'-GC**G GTA CC**C TTG TAC AGC TCG TCC ATG CC-3'), which introduced *Kpn*I restriction sites (bold), in a final volume of 50 µL containing 500 ng plasmid tpfcp/ctGFP[87], 1x Pfx buffer, 60 µM dNTPs, 0.50 µM each primer, 1.0 mM $MgSO_4$, and 1.0 U Pfx DNA polymerase (Invitrogen). The 720-bp DNA fragment was cloned into pJET1 (Fermentas) to generate pJET1/GFP. Positive clones were identified and sequenced using pJET1 and pJET2 primers. Plasmid pJET1/GFP was digested with *Kpn*I and the 720-bp insert was cloned into pTpfcp (-*Kpn*I)/tpSTK1 via a *Kpn*I site in the *tpstk1* gene to generate pTpfcp (-*Kpn*I)/ GFP-tpSTK1. The orientation of the GFP in the vector was verified by PCR (35 cycles: 94 °C 20 sec, 58 °C 20 sec, 72 °C 45 sec) using a *tpstk1* gene specific sense primer 5'-GCA TCC AAG CAG TAC ATT CC-3' and the GFP 37 gene specific antisense primer 5'- CCG TAG GTG GCA TCG CCC TCG C -3' in a final volume of 50 µL containing 500 ng pTpfcp (-*Kpn*I)/ GFP-tpSTK1, 1x PCR Buffer, 0.1 mM dNTPs, 0.50 µM each primer, and 0.25 µL Taq DNA polymerase (homemade).

The sequences of all PCR products were verified by DNA sequencing. Plasmids (5 µg) were introduced into *T. pseudonana* cells by microparticle bombardment using an established method[87]. *T. pseudonana* cells expressing $BiP_{1-27}GFP$ and $BiP_{1-39}GFP_{DDEL}$ protein fusions were provided by Dr. Nicole Poulsen.

2.23 Fluorescence Microscopy

Epifluorescence microscopy imaging was performed on an inverted microscope (Axiovert 200, Zeiss) equipped with a GFP longpass filter set (Ex: 450-490 nm, beam splitter FT 510 nm, Em: LP515; Zeiss). Confocal fluorescence microscopy was performed by Dr. Nicole Poulsen.

2.24 Fractionation of Diatom Membranes[97]

2.24.1 Differential Centrifugation

One liter of *T. pseudonana* cells (1.0-1.5 x 10^6 cells/mL) was harvested by centrifugation (3000*g*, 10 min, 18 ºC), washed once with ice cold Membrane buffer (MB) containing 250 mM sucrose and 1x protease inhibitor cocktail (Roche) (MB-SPI), and then resuspended in 3 mL of MB-SPI. The cells were equally distributed into three 2 mL tubes, 0.5 mL of glass beads were added to each tube, and the tubes were vortexed for 30 seconds. The glass beads were allowed to settle, the supernatants removed from each tube, combined, and then centrifuged at 16,000*g* for 2 minutes at 4 ºC (supernatant = fraction S0). The pellets were resuspended in 1.0 mL of MB-SPI and treated with glass beads as described above. After three lysis/centrifugation cycles, the S0 fraction was centrifuged at 16,000*g* for 15 minutes at 4 °C. The pellet (P1) was resuspended in MB-SPI and the supernatant (S1) was centrifuged at 100,000*g* for 1 hour at 4 °C yielding fractions P2 (pellet) and S2 (supernatant).

2.24.2 Sucrose Density Gradient Centrifugation

For fractionation of membranes by sucrose density gradient centrifugation the P2 fraction was brought to 20 % (w/v) sucrose in MB containing 1x protease inhibitor cocktail (Roche) (MB-PI) and layered atop a stepwise sucrose gradient (sucrose concentrations in % (w/v): 22.5, 25.0, 27.5, 30.0, 32.5, 35.0 in MB-PI). After centrifugation at 100,000*g* for 24 hours at 4 °C, each sucrose layer was collected (top layer = F1, bottom layer = F7), diluted 5-fold with MB containing 1 mM PMSF, and then centrifuged at 100,000*g* for 1 hour at 4 °C. Each membrane pellet was resuspended in 1.0 mL of MB or MB-SPI depending on application. To solubilize membranes, fractions were incubated with 1 % (v/v) Igepal in MB for 1 hour at 4 ºC, with rotation, and then centrifuged at 100,000*g* for 1 hour at 4 ºC.

2.25 Quantification of Chlorophyll[102]

Subcellular fractions in membrane buffer (MB) were extracted with acetone (80 % (v/v) final concentration) for 5 minutes on ice and then centrifuged (16,000*g*, 10 min, 4 °C). The absorbance at 652 nm of the supernatant was measured and the concentration of chlorophyll present in the supernatant was calculated from the following equation:

$$\frac{mg}{mL}\,chlorophyll = \frac{Abs_{562}\ x\ dilution\ factor}{34.5}$$

The solid residue remaining after acetone extraction was washed three times with 100 % (v/v) acetone and dried. The dried residue was dissolved in 1 % (w/v) SDS and incubated

at 95 °C for 5 minutes. Protein concentrations were determined by BCA assays (see section 2.4.1).

2.26 Precipitation of Secreted Proteins

T. pseudonana cells (density 1.0 - 1.5 x 10^6 cells/mL) were pelleted by centrifugation at 3,000*g* for 10 minutes at 18 ºC and the supernatant was filtered through a 0.2 μm filter (Corning). Proteins in the supernatant were precipitated by the addition of TCA (final concentration 8.5 % (w/v)) and sodium deoxycholate (final concentration 0.9 % (w/v)). After incubation for 15 minutes at room temperature, the sample was centrifuged at 16,000*g* for 10 minutes at 4 °C. The precipitate was washed with acetone, air-dried, dissolved in SDS sample buffer, and incubated at 95 °C for 5 minutes before SDS-PAGE[93].

2.27 Marker Enzyme Assays

The cytosolic fraction or membrane fractions from sucrose gradients (intact and detergent-solubilized) containing 10 μg total protein (as determined by BCA analysis) were used in each assay.

2.27.1 Triton-Stimulated Inositol diphosphatase (IDPase) Assay[103]

To test for the presence of Golgi membranes, subcellular fractions were incubated in assay buffer containing 30 mM Tris-MES pH 7.5, 3 mM $MgSO_4$, 50 mM KCl, 3 mM IDP, and with or without 0.025 % (v/v) Triton-X100 for 1 hour at room temperature. Proteins were precipitated as described in section 2.26. After centrifugation at 16,000*g*

for 10 minutes at 4 °C, the amount of inorganic phosphate in the supernatant was determined using the malachite green method (see section 2.20.1). The activity of IDPase (nmol P_i released/min) was obtained by the difference of activity of samples with Triton and those without Triton.

2.27.2 Vanadate sensitive adenosine triphosphatase (vsATPase) Assay[104]

To test for the presence of plasma membranes, the subcellular fractions were incubated in assay buffer containing 30 mM Tris-MES pH 6.5, 3 mM $MgSO_4$, 50 mM KCl, 8 mM ATP, 100 mM KNO_3, 5 mM NaN_3, and with or without 0.025 % (v/v) Triton-X100 for 1 hour at room temperature. To identify ATPases sensitive to vanadate, 50 μM Na_3VO_4 was added to the assay buffers. Proteins were precipitated as described in section 2.26 and the amount of inorganic phosphate in the supernatant was determined using the malachite green method (see section 2.20.1). The activity of vsATPase (nmol P_i released/min) is the loss of activity in the presence of vanadate from the total ATPase activity (i.e. without vanadate added).

2.27.3 Antimycin insensitive NADH-cytochrome *c* reductase (CCRase) Assay[103]

To test for the presence of ER membranes, subcellular fractions were added to assay buffer containing 50 mM sodium phosphate pH 7.5, 30 μM cytochrome *c*, and with or without 1 μM antimycin A. The reaction was initiated by the addition of 0.1 mM NADH. The change in absorbance was measured at 550 nm (ε_{550}=18.5 mM^{-1} cm^{-1} for cyt *c*). The activity of CCRase (nmol Cyt *c* reduced/min) is obtained by the difference of activities in samples with antimycin A from those without antimycin A.

2.27.4 Cytochrome c oxidase (COx) Assay[103]

To test for the presence of mitochondria subcellular fractions were added to assay buffer containing 50 mM sodium phosphate pH 7.5, 5 mM $MgCl_2$, and 10 mM KCl, and with or without 0.02 % (v/v) Triton-X 100. Cytochrome *c* was reduced by molar excess of sodium dithionite and then aerated with nitrous gas for 5 minutes to remove excess sodium dithionite. The reaction was initiated by the addition of 80 µM cyt *c* (reduced) and the change in absorbance was measured at 550 nm (ε_{550}=18.5 mM^{-1} cm^{-1}).

2.28 Protease Accessibility Assay[97]

Intact and detergent solubilized (1 % (v/v) Igepal) membranes from fractions F4 and F5 (10 µg of total protein) in MB containing 250 mM sucrose were incubated in the absence or presence of chymotrypsin (0.3 µg/µL or 1.0 µg/µL) for 1 hour on ice. To inhibit the protease, PMSF was added at a final concentration of 5 mM. For SDS-PAGE and Western blot analysis (see section 2.15), an equal volume of 2x SDS loading buffer was added to the membrane fractions and incubated at 95 °C for 15 minutes. In addition to Western blot analysis, the membranes were assayed for CCRase activity or IDPase activity (according to 2.27.3 and 2.27.1, respectively).

2.29 Antibody Inhibition Assays[97]

T. pseudonana cellular membrane fractions F4 and F5 (intact and detergent-solubilized) were incubated at 4 °C for 2 hours with the indicated concentrations of IgG or in the absence of IgG. Subsequently, luminescence-based kinase assays using membrane fractions containing 10 µg of total protein were performed as described in section 2.16.2.

CHAPTER 3: RESULTS

3.1 Recombinant silaffins

3.1.1 Purification of recombinant silaffins[92]

To be tested as substrates in kinase assays, four recombinant silaffin domains (termed rSil3, rSil1L, rSilN, and rSilC) encoded by select regions of *Thalassiosira pseudonana* (*tpsil3*, *tpsil1l*) or *Cylindrotheca fusiformis* (*sil1*) silaffin genes were expressed in *E. coli*, with each recombinant silaffin possessing a C-terminal hexahistidine tag (Appendix A). Characteristics of these proteins are shown in Table 3.1. One feature to note is the abundance (21 % to 43 % of the total protein) of hydroxyl amino acids, especially serine and threonine, in the recombinant silaffins that are potentially phosphorylated by tpAK1 or tpSTK1.

Table 3.1. Properties of recombinant silaffins. The theoretical molecular mass (M_{theor}) in kilodaltons (kDa) and theoretical isoelectric point (pI) of each recombinant silaffin are based on the amino acid sequence of the protein without any modifications. S+T = content of serine and threonine, respectively. T = content of tyrosine.

	rSilC	**rSil1L**	**rSil3**	**rSilN**
M_{theor} (kDa)	17.6	11.8	22.1	9.9
pI_{theor}	11.8	10.6	9.4	3.8
S + T	36 %	23 %	21 %	30 %
Y	7.3 %	1.8 %	0 %	0 %

The presence of a C-terminal hexahistidine tag allowed for purification of the recombinant silaffins to homogeneity via Ni^{2+}-NTA chromatography. SDS-PAGE analysis demonstrated the purity of the recombinant silaffins (Figure 3.1). However, the recombinant silaffins exhibited lower mobilities (up to ~50 % lower) through an SDS-PAG than predicted by their theoretical molecular masses. The discrepancy in mobilities may be that the hydrophilic nature of recombinant silaffins results in less binding of SDS.

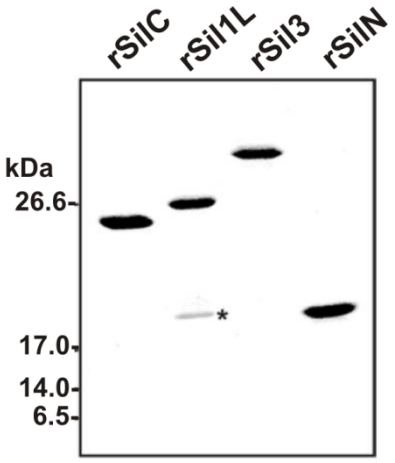


Figure 3.1. Recombinant silaffins rSilC, rSil1L, rSil3, and rSilN. Coomassie-stained SDS-PAG of recombinant silaffins purified via Ni^{2+}-NTA chromatography. The band in the rSil1L lane labeled with an asterisk represents the monomer, whereas the band of higher molecular mass corresponds to an unusually stable dimer.

Delayed migration of proteins in an SDS-PAG is also observed when the proteins carry posttranslational modifications. In theory, this should not be the case for the recombinant silaffins as *E. coli* does not incorporate posttranslational modifications. Indeed, mass spectroscopic analysis of the recombinant silaffins confirmed that these proteins carry no posttranslational modifications (see Appendix B for mass spectra).

3.1.2 Quantification of recombinant silaffins

Colorimetric methods, like the bicinchoninic acid (BCA) assay[94], are commonly used to determine protein concentration because these methods are fast (results within minutes) and inexpensive. However, BCA color formation is strongly influenced by the presence of cysteine, tyrosine, and tryptophan residues, often leading to an overestimation of the protein concentration[105]. Although more tedious, amino acid analysis (AAA) is less sensitive to the amino acid composition of a protein and, therefore more accurately determines the protein concentration[106]. To determine the amount of recombinant silaffins purified from *E. coli*, both amino acid analysis and BCA assays were performed. The concentrations of each protein determined by AAA and BCA analysis are shown in Table 3.2.

Table 3.2. Protein concentrations of recombinant silaffins determined by Amino Acid analysis (AAA) and BCA analysis (BCA).

	rSilC	**rSil1L**	**rSil3**	**rSilN**
BCA (mg/mL)	10.0	3.3	15.8	10.0
AAA (mg/mL)	2.9	1.0	7.1	2.8

The BCA protein concentrations for all of the recombinant silaffins were significantly higher (at least 2-fold) than the concentrations obtained by amino acid analysis. For three of the recombinant silaffins (rSilC, rSil1L, and rSil3) the overestimation of the BCA assay can, in part, be attributed to the presence of cysteine

and tyrosine residues as these amino acids can form colored, copper-BCA complexes independent of the protein. Recombinant silaffin rSilN however, does not contain any of these residues and its concentration was the most overestimated (3.5-fold) of all of the proteins tested.

3.1.3 Mineral forming capabilities of recombinant silaffins[92]

To determine if the recombinant silaffins retained their silica precipitating ability, each recombinant silaffin was incubated for 10 minutes in silicic acid solutions covering the pH range of 4-10. At pH values in the range of 6-10, recombinant silaffins rSil3, rSil1L, and rSilC exhibited precipitation activity in a dose-dependent manner (i.e. the amounts of silica produced increased linearly with the amounts of silaffin added to the silicic acid solutions (Figure 3.2)). As rSilN (the only negatively charged silaffin) and the control without silaffins did not induce silica precipitation, the formation of these silica precipitates is presumably due the interaction of the polycationic silaffins (rSil3, rSil1L, and rSilC) with the negatively charged colloidal polysilicic acid particles that form rapidly at pH $\geq$ 6.0. Although the recombinant silaffins lack posttranslational modifications, these proteins are capable of forming biominerals *in vitro* and are therefore good substrates to test the phosphorylation activities of putative kinases tpAK1 and tpSTK1.

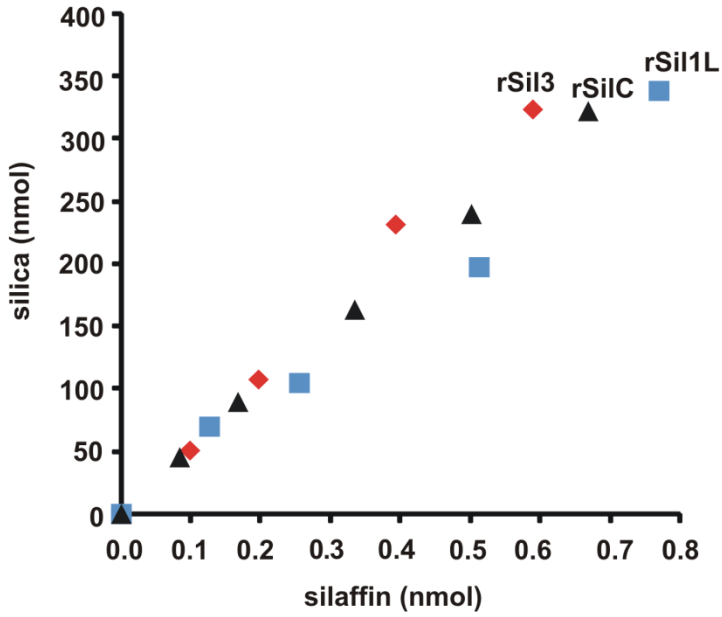


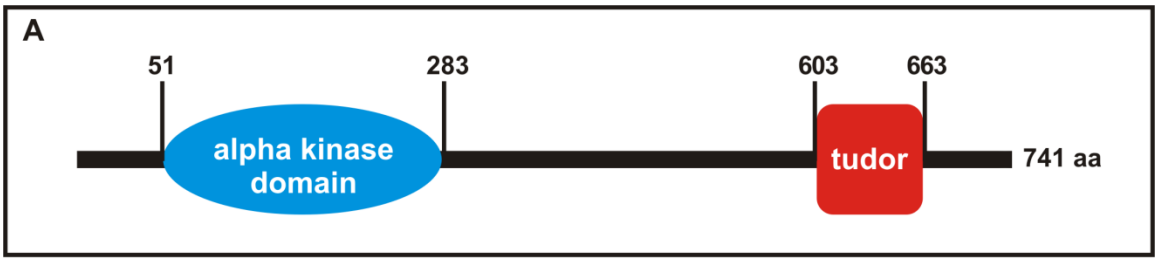
Figure 3.2. Silica formation by recombinant silaffins. Correlation between the amounts of recombinant silaffin added to a buffered silicic acid solution (50 mM sodium phosphate/citrate pH 7) and the amount of silica precipitated. Each data point represents the average of three independent measurements (standard deviation ≤ 10 %). The amounts of silica produced were determined using the β-silicomolybdate method[95].

3.2 Putative kinase tpAK1

3.2.1 TpAK1 DNA sequence analysis

TpAK1 was one of two putative kinases identified in a combined proteomics and genomics study by Hildebrand and coworkers[65] (see section 1.5). Reverse transcription PCR (RT-PCR) and RACE PCR showed that the *tpak1* gene (see Appendix C) contains only four introns instead of the five predicted by the *tpak1* gene model in the *Thalassiosira pseudonana* genome database[85]. The *tpak1* gene encodes a 741 amino acid protein (83 kDa) with an isoelectric point of 6.14 (Figure 3.3 A). Sequence analysis using

the BLAST[107] algorithm revealed that the N-terminal domain exhibits homology (e-value: e-29) to the catalytic domain of alpha kinases, a unique family of eukaryotic kinases that share no sequence homology with other kinase families (Figure 3.3 B). Although alpha kinases have no detectable sequence identity with the other kinase families, the alpha kinase domain is structurally similar to serine/threonine kinases and contains many of the conserved residues required for protein phosphorylation[108]. TpAK1 is predicted, by sequence analysis programs[108], to contain a Tudor domain whose general function is currently unknown. No N-terminal signal peptide for targeting to the endoplasmic reticulum (ER) (based on analysis using Signal P program[109]) or transmembrane domains are predicted.



B

```
        10         20         30         40         50         60
MTATPYNREE VQRSLREKIW RAAATAVSEE DPWKKHDITS IPAERVVRHC YNPNTRQFKK

        70         80         90        100        110        120
DETIVKVEKE PFTHGAMRHC FRLKKLATPP QSSSNHRFHS YGWSRALNYV AKCYLKEDGT

       130        140        150        160        170        180
IDTSRKGTDN VLTDIILQYE AAHWSALFNE ANPPRKIDFI RAYAMEFVDR PGKPMFAVER

       190        200        210        220        230        240
YIAGNDSYGC GFHKHNTNSG FIDLEIRRKT PQVFSAHSFY ASEGTRLVAD VQGVGDLYTD

       250        260        270        280        290        300
PQVLSIDYRF GDGDLGPRGM ALFFKTFRHC DMSDLLGIPI FPLSRNERRH QAKYDDEEST

       310        320        330        340        350        360
LSEASSAVEE DLRCRFKMLD ANRQRRKTVL MRPIDIMQSE DKADTAKRSN ISNVSLTIRQ

       370        380        390        400        410        420
SMRQLNIQKS VIHRSKSDVD EITSCLQMAV TDAVFDHTAF HRYESGELKP RHFHGTPKHQ

       430        440        450        460        470        480
TQLTVVPVKA SPPMKITDET KANLGKVHYH LACLHGLDRF PEIVPTNASS GVEDLPSHDV

       490        500        510        520        530        540
FSVVFHLSHA ASLYNVPACL SLARALIGLD SSVSPLLKAN VRIDFELSKD LCWRGMTAQK

       550        560        570        580        590        600
ARAAPKVAAG CLLYQILEDE GTTGDVEKMN ILEETLNQMK VATEEEAILK EHANKLTRGK

       610        620        630        640        650        660
AAGFHVGDKV EGNYFMEGTF YSGVVVEVLD GGNSVVVQYD DDGSAEQLTI DNVRSIEPAT

       670        680        690        700        710        720
AITDDQNSPL SDEEALGSVN TDEECLFEDY ELMAKLADLK VKAGKRNDAR VLYQDAAELA

       730        740
MSAGKMKTAN CLSMKAAQLE G
```

Figure 3.3. Putative kinase tpAK1. ***A,*** Schematic representation of predicted tpAK1 domain structure. The blue oval represents the predicted alpha kinase domain and the orange box indicates the tudor domain. Numbers indicate the amino acid positions at the borders of the domains. ***B***, Primary amino acid sequence of tpAK1.

3.2.2 Recombinant tpAK1 Substrate Specificity

To investigate whether tpAK1 is a protein kinase, the full length protein was expressed as a fusion with a C-terminal hexahistidine tag (see Appendix D) and purified from *E. coli* (Figure 3.4). TpAK1 was insoluble in *E. coli* (inclusion bodies) and therefore had to be purified under denaturing conditions (8 M urea). The purified protein was subjected to conditions that promote protein refolding (see section 2.10.2 for experimental details).

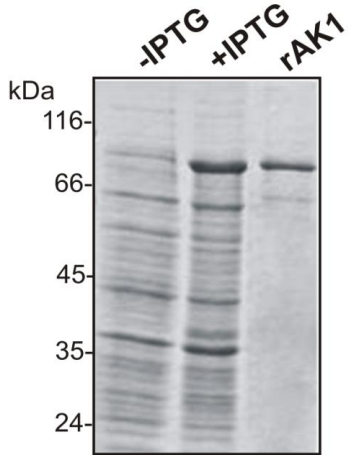


Figure 3.4. Expression and Isolation of recombinant tpAK1. Coomassie-stained SDS-PAG. Identical amounts of cells from an *E. coli* clone harboring a pET11a plasmid encoding recombinant tpAK1 were loaded before (-) and after 3 hours of induction with IPTG (+). After solubilization of the inclusion bodies with urea and adsorption of the solubilized material to Ni-NTA agarose, pure recombinant tpAK1 (rAK1) was refolded by dialysis.

To determine whether tpAK1 is a protein kinase, its ability to phosphorylate purified recombinant silaffins and commercial kinase substrates (dephosphorylated casein, myelin basic protein, histone I, and histone II) was tested. In kinase assays using

gamma phosphate radiolabeled ATP ([γ-^{32}P]-ATP) as the phosphate donor, the incorporation of radiolabeled phosphates to the protein substrates by tpAK1 was analyzed by SDS-PAGE. Recombinant tpAK1 showed no phosphorylation activity with any of the substrates tested (Appendix E). The lack of kinase activity may be due to:

i) Recombinant tpAK1 was not folded into an active conformation under the applied refolding conditions.

ii) Recombinant tpAK1 has substrate specificity for substrates different from those that were tested.

iii) Recombinant tpAK1 may require an unknown cofactor for its activity.

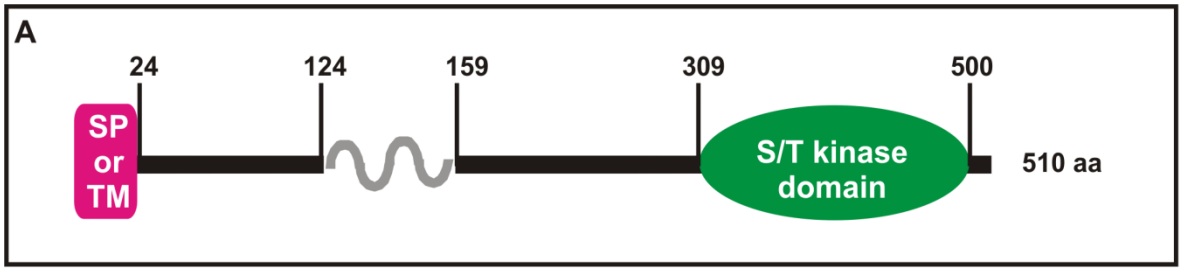
Due to the lack of phosphorylation activity, further characterization of tpAK1 was discontinued.

3.3 Putative kinase tpSTK1

3.3.1 TpSTK1 DNA sequence analysis

The full length *tpstk1* gene, the other putative kinase from the Hildebrand study[65], was determined by RACE-PCR and RT-PCR. RT PCR confirmed the *tpstk1* gene model in the *Thalassiosira pseudonana* genome database (see Appendix F). Analysis of the 55 kDa protein (Figure 3.5 B) encoded by the *tpstk1* gene using primary sequence analysis programs[107-108] revealed a predicted a coiled-coil domain and an ATP binding kinase domain with homology (e value: e-16) to the catalytic domain of serine/threonine kinases (Figure 3.5 A). The first 24 amino acids are predicted to be an N-terminal signal sequence for co-translational import into the ER or a permanent membrane anchor, as the

structural characteristics of both sequence elements are very similar, making them indistinguishable from one another[110].



B

```
        10         20         30         40         50         60
MRVIRNVSLV VIVTSCTFLV SVTSFTSQRT VTSASPSYTN SNNNLYSPPL FASGPPVISN

        70         80         90        100        110        120
WRYQSSSGTI SGTVTNHPTI PDGDSITTSP VRDGNGINEN TIVQTKSGSK YKLSNVAWGA

       130        140        150        160        170        180
TIPAVETKKA MQKEAEAAEK ARKVEAEKAK TAAAKAKTAS KATPKATSPA AATTKQPSSA

       190        200        210        220        230        240
ELRRLAKSTF SLTGNTVGSN GKYLLAGKPR QSSGRAAKIW TAYIADPIEP GVPLGFSEGD

       250        260        270        280        290        300
ESKVQPLTIK LSPDVDRLYL ENANYNKVQS GLFMGRFVKK MDYIENVPSS DRSLDGKVSA

       310        320        330        340        350        360
LIIESGQYDL KALLSARKGL PLAGKALRDA AAAAGQCIQA VHSSNLVWTD LKTENFIVVR

       370        380        390        400        410        420
NGNLYQEVED VKGSMGLPGV KGIDLESVIA KGGNPIDYSP EACPPEFAKA YMEGIGGDFV

       430        440        450        460        470        480
LDYSYDMWSL GMMLYELSTG KAYFDKKGPS QVTKLLCTDT FEADVSAVPD EKLRDLIGNC

       490        500        510
LNLDPRKRPD ITGFLLHPYF LLSGIGPISF
```

Figure 3.5. Putative kinase tpSTK1. ***A,*** Schematic representation of the predicted tpSTK1 domain structure. SP or TM = putative signal peptide or transmembrane domain, respectively. The gray zig-zig indicates the predicted coiled-coil, and the red oval shows the putative serine threonine (S/T) kinase domain. Numbers indicate the amino acid positions at the borders of the domains. ***B***, Primary amino acid sequence of tpSTK1.

3.3.2 Cell cycle specific expression of tpSTK1

As observed by Hildebrand and co-workers[65], the expression of *tpstk1* mRNA is strongly up-regulated during silica formation in synchronized *T. pseudonana* cells, which closely follows the expression pattern of the silaffin-encoding gene, *tpsil3,* and is similar to that of *tpsil1*, another silaffin encoding gene (Figure 3.6). The increased *tpsil1* mRNA expression one hour into the cell cycle progression coincides with the production of a few girdle bands at this time. Given that silacidins are also embedded in the diatom biosilica and are able to precipitate silica *in vitro*[62], the mRNA expression of the silacidin encoding gene *tpsic* from 0 hours to 10 hours in synchronized *T. pseudonana* cells was monitored. As observed with the silaffins, the mRNA expression of the silacidin-encoding gene, *tpsic* is upregulated during valve formation (Figure 3.6). However, the expression levels during valve formation are relatively low when compared to the silaffins.

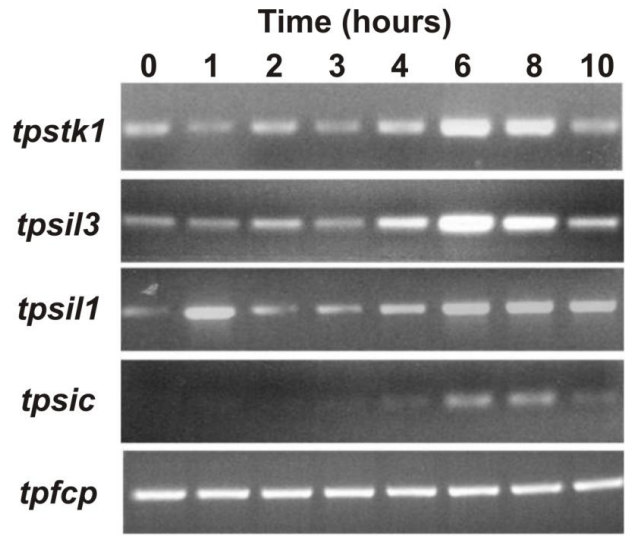


Figure 3.6. Cell cycle-specific mRNA expression of native tpSTK1. Time course of *tpstk1* mRNA expression in synchronized *T. pseudonana* cells. The times refer to hours after addition of silicic acid to silicon-starved cells. For each time point, cells from equal culture volumes were used for cDNA synthesis. Reverse transcription was performed from identical amounts of cDNA using primers specific for the indicated genes: *tpsil3* and *tpsil1*, genes encoding silaffins tpSil3 and tpSil1, respectively; *tpsic*, gene encoding silacidin tpSic; *tpfcp*, gene encoding a fucoxanthin chlorophyll-associated protein (tpFCP) that is constitutively expressed in *T. pseudonana* under constant illumination (A. Scheffel and N. Kröger, unpublished communication).

To investigate expression of the tpSTK1 protein during the *T. pseudonana* cell cycle, a polyclonal antiserum was produced against a recombinant version of the protein (see section 2.14). From the antiserum, anti-tpSTK1 IgG antibodies were obtained by dual affinity chromatography on a recombinant tpSTK1-loaded column followed by protein G-agarose. The anti-tpSTK1 IgGs specifically detected a protein with an apparent molecular mass of about 60 kDa in a *T. pseudonana* whole cell lysate (Figure 3.7 A), which closely matches the predicted molecular mass of tpSTK1 (55 kDa). Western blot analysis of whole cell lysates from a synchronized *T. pseudonana* culture detected an increase in tpSTK1 protein expression beginning at 6 hours (valve formation is at a

maximum between 4 and 6 hr), demonstrating that biosynthesis of the tpSTK1 protein closely followed expression of the mRNA (Figure 3.7 B). The high level of the tpSTK1 protein was maintained at 8 and 10 hours, whereas the *tpstk1* mRNA level decreased over the same period of time.

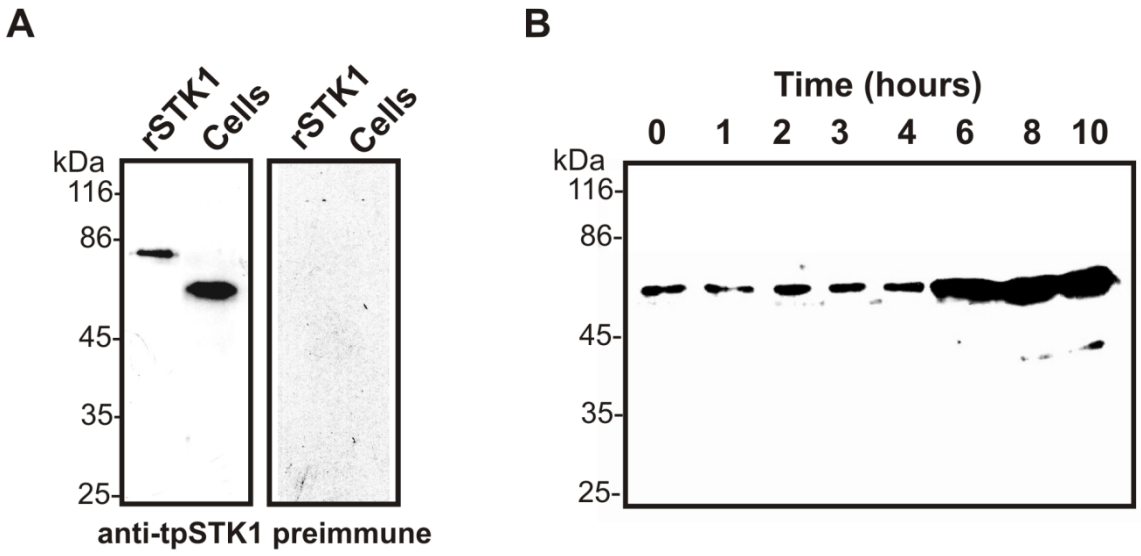


Figure 3.7. Cell cycle-specific protein expression of native tpSTK1[97]. ***A***, Western blots probed with anti-tpSTK1 IgG (0.2 µg/mL) and preimmune IgG (0.2 µg/mL) that were isolated from the same rabbit. Lanes "rSTK1" contained 100 ng of purified recombinant tpSTK1; lanes labeled "cells" contained lysates from 1.3×10^6 *T. pseudonana* cells. ***B***, Time course of tpSTK1 protein expression in synchronized *T. pseudonana* cells. The times refer to hours after addition of silicic acid to silicon-starved cells. For each time point, cells from equal culture volumes were used for Western blot analysis probed with 0.2 µg/mL anti-tpSTK1 IgG.

3.3.3 Recombinant tpSTK1 Substrate Specificity

A recombinant tpstk1 gene was constructed that encoded a histidine-tagged derivative of tpSTK1. Recombinant tpSTK1 (Appendix G) was expressed in *E. coli*, solubilized from inclusion bodies (using 8 M urea), and purified to homogeneity by immobilized metal affinity chromatography (Figure 3.8). The purified protein was subjected to conditions that promote protein refolding (see section 2.13.2 for experimental details).

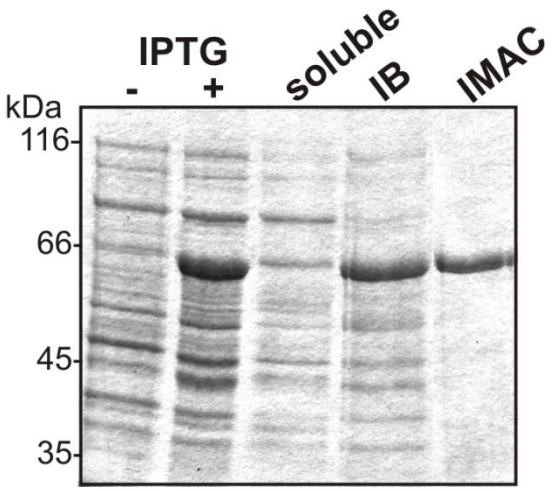


Figure 3.8. Expression and isolation of recombinant tpSTK1[97]. Coomassie-stained SDS-PAG. Identical amounts of cells from an *E. coli* clone harboring a pPROEX-Htc plasmid encoding recombinant tpSTK1 were loaded before (-) and after 3 hours of induction with IPTG (+). After cell lysis equal aliquots of the 23,000*g* supernatant (soluble) and pellet (inclusion bodies = IB) were loaded. After solubilization of the inclusion bodies with urea, adsorption of the solubilize material to Ni-NTA agarose, and refolding on the column, the majority of recombinant tpSTK1 eluted with 200 mM imidazole (immobilized metal affinity chromatography = IMAC).

Radioactive kinase assays in the presence of [γ-^{32}P]-ATP revealed that recombinant tpSTK1 phosphorylates three of the four recombinant silaffins (rSil3, rSil1L, and rSilC) tested (Figure 3.9 A). Recombinant silaffin rSilN was not phosphorylated by recombinant tpSTK1, and is the only recombinant silaffin tested to possess an acidic isoelectric point (pI = 3) while those of the phosphorylated recombinant silaffins range from 8-12 (see Figure 3.9 B). Of the common kinase substrates tested (myelin basic protein: MBP, histones I and II: HisI and HisII, dephosphorylated casein: d. Cas.)[111], the only one not phosphorylated by recombinant tpSTK1 was dephosphorylated casein which, like rSilN, has an acidic isoelectric point (pI = 3.5) whereas the other commercial kinase substrates have isoelectric points between 10 and 12. Bovine serum albumin (BSA) and lysozyme were also tested in the assays but neither substrate was phosphorylated by tpSTK1 even though lysozyme has a basic isoelectric point of 11.4. These results indicate that tpSTK1 strongly prefers substrates with basic isoelectric points (pI > 7). However, pI appears not to be the only criterion for phosphorylation by tpSTK1, as demonstrated by the lack of activity with lysozyme.

To provide a quantifiable, rapid, non-radioactive method to detect the phosphorylation activity of recombinant tpSTK1, a luminescence-based kinase assay that measures ATP consumption was employed[99]. The data obtained from the luminescence assays are analogous to the information obtained with the radioactive assays (Figure 3.9 B).

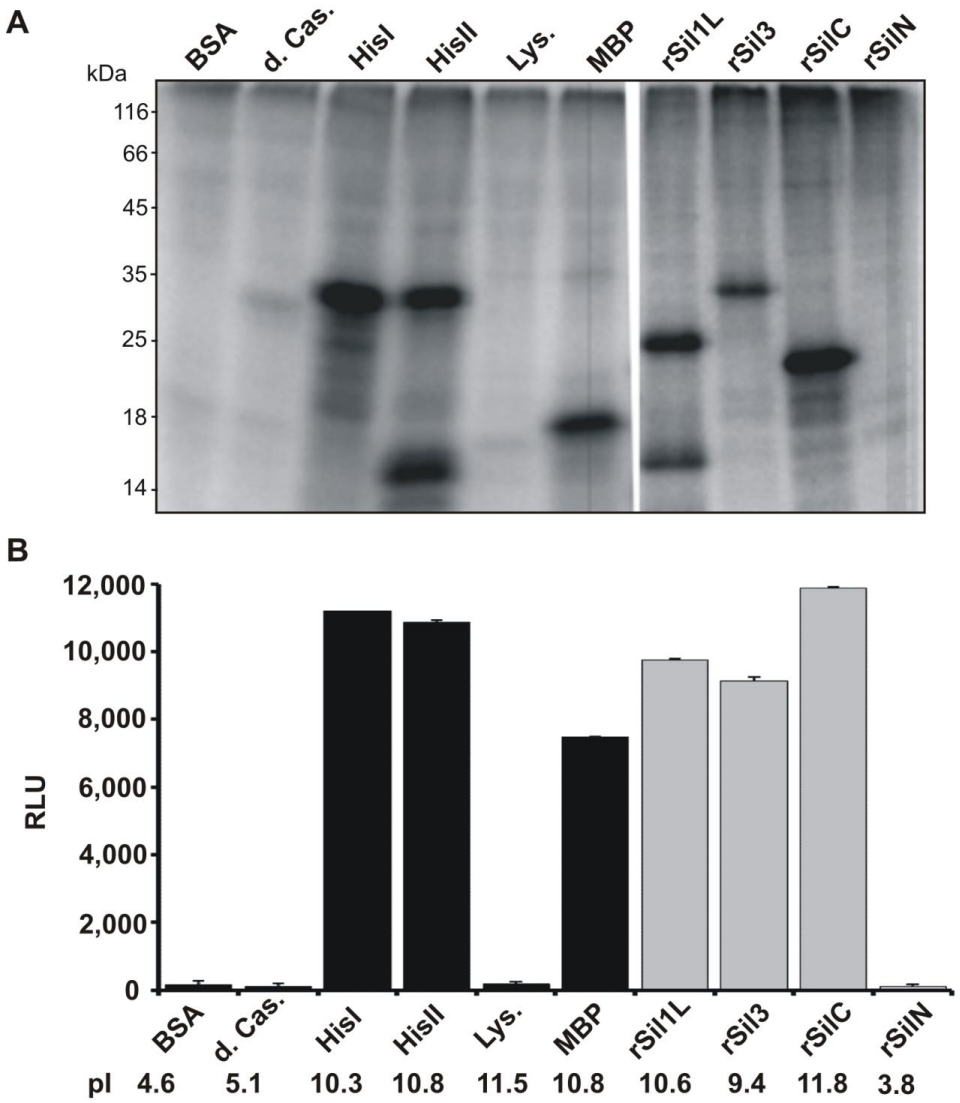


Figure 3.9. Phosphorylation activity of recombinant tpSTK1 with recombinant silaffins (rSilN, rSilC, rSil3, and rSil1L) and commercial proteins as substrates. ***A***, SDS-PAGE analysis of recombinant tpSTK1 phosphorylation activity in the presence of 2 µCi [γ-^{32}P] ATP and 0.10 mg/mL substrate protein. ***B***, Luminescence-based kinase assay[99] with 0.10 mg/mL substrate protein. Each measurement was repeated at least three times. RLU, relative luminescence units. Lys- Lysozyme, His-Histone, MBP- Myelin Basic Protein, d. Cas- dephosphorylated casein, BSA- bovine serum albumin. The isoelectric point of each protein is indicated below the protein name.

3.3.3.1 Recombinant silacidin substrates

In addition to the recombinant silaffin and commercial protein substrates, silacidin substrates were tested in the luminescence-based kinase assays using recombinant tpSTK1. Silacidin[62] had not been identified when performing the radioactive kinase assays that included recombinant tpAK1; therefore, the phosphorylation activity of recombinant tpAK1 with silacidin as the substrate has not been tested. The first silacidin substrate tested for phosphorylation by recombinant tpSTK1 was a synthetic 19-mer peptide (SDESEESEDSVSSEDEDWW; pI 2.8) corresponding to a 17 amino acid fragment of silacidin A (see Appendix H). As a positive control, a 19-mer fragment derived from the silaffin gene *sil1,* R5 was utilized in the kinase assays because of its basic isoelectric point (pI 11.6) and the fact that rSilC (a recombinant protein derived from the same gene) was phosphorylated. The activity of recombinant tpSTK1 was negligible for both R5 and Sic1 peptides at the standard substrate concentration of 10 μg. However, at 10-fold increased substrate concentrations, recombinant tpSTK1 phosphorylation activity on R5 was comparable to activities with recombinant silaffin rSilC as a substrate (Figure 3.10 black bars), whereas Sic1 was not phosphorylated. These results are consistent with the preference of tpSTK1 for basic substrates but it is also revealed that proteins rather than peptides are substrates for recombinant tpSTK1.

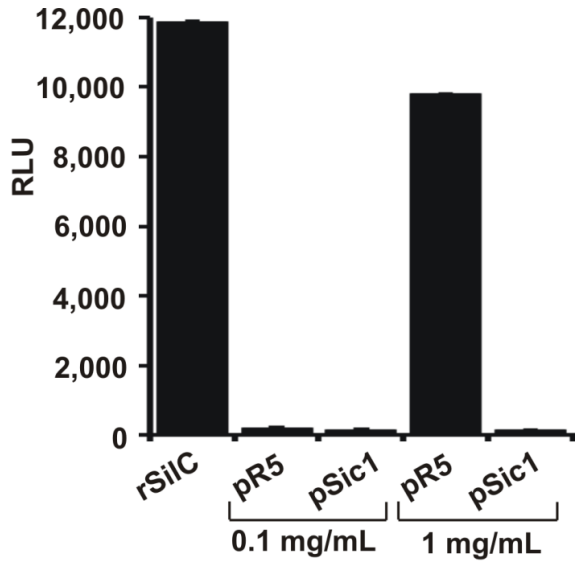


Figure 3.10. Phosphorylation activity of recombinant tpSTK1 with synthetic peptides (R5 and Sic1) and recombinant silacidin substrates. Kinase activity with silaffin peptide R5 (pR5) and silacidin peptide Sic1 (pSic1) at normal (0.1 mg/mL) and tenfold increased (1 mg/mL) substrate concentration. rSilC serves a reference for tpSTK1 kinase activity.

Since recombinant tpSTK1 prefers protein substrates, efforts were made to purify a full length silacidin protein that could be used in tpSTK1 kinase assays. The first silacidin construct that was designed encoded the mature silacidin protein with a C-terminal hexahistidine tag. However, expression of the protein in *E. coli* was not observed. The lack of silacidin expression may be attributed to poor, undetectable yields of the protein being produced by the cells. One method to improve the yield of recombinant protein expression is to add an N-terminal tag (e.g. glutathione S-transferase and maltose-binding protein)[112]. Since affinity tags, such as those mentioned above, could potentially interfere with the functional studies of tpSTK1, recombinant silaffin rSilN was chosen as the N-terminal tag for the following. 1) rSilN is a small protein, only 9.9

kDa. 2) Expression of rSilN in *E. coli* has been achieved. 3) There are no additional amino acids (called a linker region) between the cleavage site and rSic that could remain on the N-terminus of rSic. 4) rSilN was not phosphorylated by recombinant tpSTK1 and therefore should not contribute to any observed phosphorylation activity on the fusion protein. Recombinant protein $rSilN_{H10}$-rSic (with the N-terminal protein rSilN possessing a C-terminal decahistidine tag, see Appendix I) was expressed and purified from *E. coli* (Figure 3.11 A and B). Due to the presence of the methionine, the mature recombinant silacidin protein carrying no additional amino acids can be cleaved from $rSilN_{H10}$ with cyanogen bromide (which cleaves on the carboxy side of methionine residues) (Figure 3.11 A). To separate rSic from $rSilN_{H10}$ and from uncleaved or partially cleaved $rSilN_{H10}$-rSic, immobilized metal affinity chromatography was employed yielding pure rSic in the flow through fraction (FT) as rSic lacks a histidine tag and cannot bind to a Ni-NTA column. In the luminescence based kinase assays, neither $rSilN_{H10}$-rSic nor the purified rSic was phosphorylated by tpSTK1, likely due to the acidity of the substrates (pI = 3.8 and 3.6, respectively) (Figure 3.11 C).

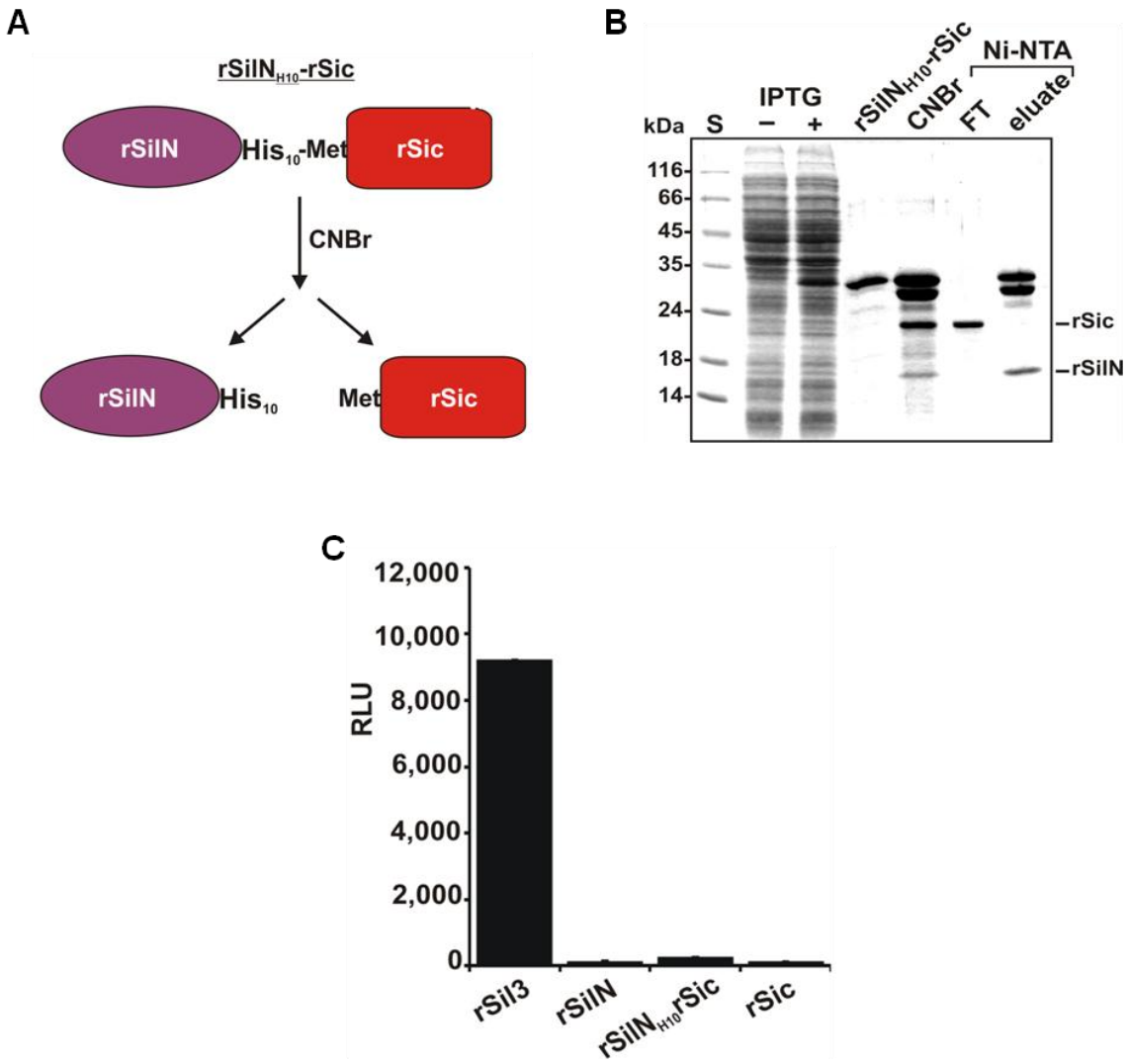


Figure 3.11. Phosphorylation activity of tpSTK1 with recombinant silacidin substrates rSilN$_{H10}$rSic and rSic. ***A,*** Schematic of cleavage of rSil$_{N10}$rSic with cyanogen bromide (CNBr) to yield the individual proteins. ***B,*** Coomassie-stained SDS-PAG. **S**: molecular mass standard. *IPTG***:** Identical amounts of cells from an *E. coli* clone harboring plasmid pET28a/rSilN$_{H10}$-rSic were loaded before (-) and after 3 hours (+) of induction with IPTG. *rSilN$_{H10}$-rSic*: Silaffin-silacidin fusion protein after purification by IMAC. *CNBr***:** Silaffin-silacidin fusion protein after CNBr treatment. *Ni-NTA***:** CNBr fragments of rSilN$_{H10}$-rSic that did not bind (*FT*) and did bind (*eluate*) to Ni-NTA resin. ***C,*** Luminescence-based kinase assay with 0.10 mg/mL substrate protein. Each measurement was repeated at least three times. RLU, relative luminescence units. rSil3 and rSilN were added a reference of tpSTK1 kinase activity.

3.3.3.2 Hydrogen fluoride treated native silaffin substrates

As silaffins possess numerous posttranslational modifications, the enzymes involved in applying the modifications are likely synergistic (i.e. the action of one enzyme may promote or enhance the activity of another enzyme). To test whether the phosphorylation of silaffins by tpSTK1 is enhanced by the presence of a posttranslational modification, hydrogen fluoride (HF) treated native silaffins (tpSil3-HF, tpSil1/2L-HF, tpSil1/2H-HF) were used in kinase assays. HF treatment removes *O*-linked posttranslational modifications such as *O*-linked glycosylation and phosphate moieties[113]. The HF treated silaffins still contain their modified alkylated lysine residues that are resistant to HF[58-59,61]. In the luminescence-based kinase assays phosphorylation of the HF treated silaffins by recombinant tpSTK1 was observed, albeit to a significantly lesser extent than the phosphorylation of recombinant silaffins rSil1L and rSil3. (Figure 3.12). This result showed that recombinant tpSTK1 phosphorylation is hindered by the presence of the modified alkylated lysine residues. This may be evidence that protein phosphorylation by tpSTK1 occurs before alkylation *in vivo*.

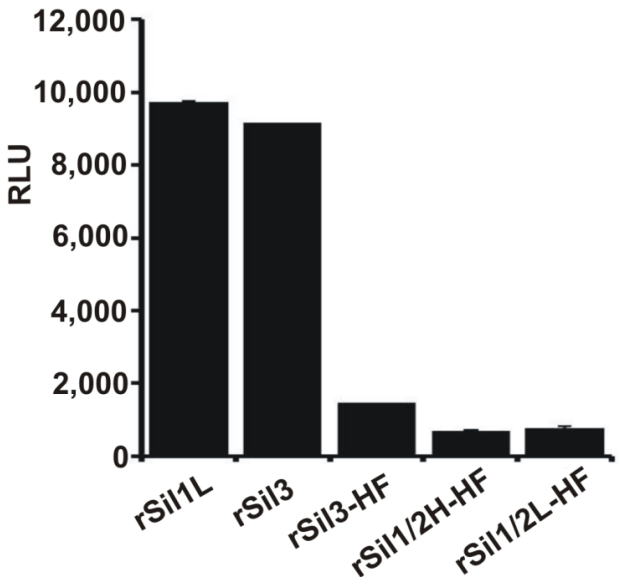


Figure 3.12.Recombinant tpSTK1 phosphorylation activity with HF-treated native silaffins. Kinase assays with 0.1 mg/mL *T. pseudonana* HF-treated silaffins (Sil3-, Sil1/2, Sil1/2-HF) contain posttranslational modifications and the activity of tpSTK1 on these substrates was compared to the activities with recombinant silaffins rSil1L and rSil3. HF treated silaffins still contain modified lysine residues. Each measurement was repeated at least three times. RLU, relative luminescence units.

3.3.4 Requirements for recombinant tpSTK1 activity

3.3.4.1 Co-substrates and co-factors

Recombinant tpSTK1 utilizes ATP for its phosphorylation activity. However, kinases, such as casein kinase II, can also hydrolyze GTP[68,79]. To test if recombinant tpSTK1 utilizes GTP as its co-substrate, GTP was added to the kinase reaction mixture at different ratios to ATP concentration. Since the luminescence assay only measures ATP consumption, a decrease in relative luminescence in the presence of GTP would be indicative of the kinase being able to also hydrolyze GTP. With rSil3 as the substrate the kinase activity remains unchanged at all GTP:ATP ratios tested, demonstrating that recombinant tpSTK1 does not utilize GTP (Figure 3.13 A). Divalent metal cations,

(usually Mg^{2+}) are typically used as co-factors by kinases in protein phosphorylation[68]. Recombinant tpSTK1 requires a divalent cation co-factor for phosphorylation of recombinant silaffins (Figure 3.13 B). To fulfill the divalent cation requirement for its activity, recombinant tpSTK1 preferred Mg^{2+} to any of the other divalent cations tested in the kinase assays (Figure 3.13 B). Negligible amounts of kinase activity were detected in the presence of Ca^{2+} or Zn^{2+} (in the luminescence based assays). In the presence of Mn^{2+}, recombinant tpSTK1 is moderately active, exhibiting nearly 50 % of the full phosphorylation activity on rSil3.

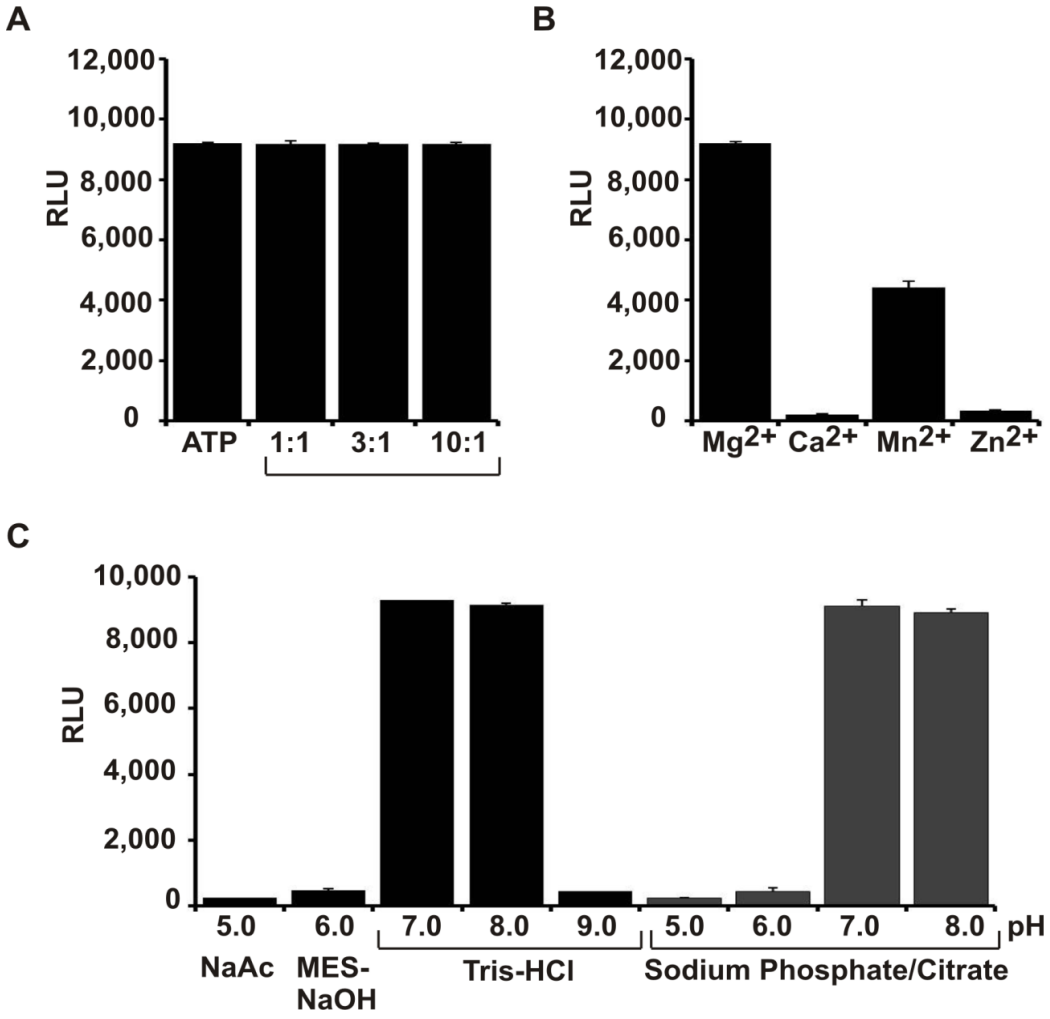


Figure 3.13. Co-factor dependence and pH stability of recombinant tpSTK1[97]. ***A***, Kinase activity with rSil3 in the presence of ATP and increasing concentrations of GTP (molar ratios of the nucleotides are indicated). The consumption of ATP was not affected by the presence of GTP indicating that GTP cannot replace ATP as a co-substrate for recombinant tpSTK1 (note: GTP cannot substitute for ATP in the luciferase-dependent reporter reaction[99]). ***B***, Metal-ion dependence of kinase activity. After refolding, recombinant tpSTK1 was incubated under kinase reaction conditions in the absence or presence of 10 mM of the indicated cation (no other polyvalent metal ions were present). ***C***, Kinase activity in different buffers ranging from pH 5.0 – 9.0. The concentration of Mg^{2+} in the sodium phosphate citrate buffers was 10x normal concentration to yield comparable activities. Each measurement was repeated at least three times. RLU, relative luminescence units. NaAc = sodium acetate.

3.3.4.2 pH dependence

As silaffin phosphorylation takes place within the endomembrane system of diatoms, the ability of recombinant tpSTK1 to phosphorylate rSil3 was tested at different pH levels that correspond to the known pH inside different intracellular compartments (i.e. endoplasmic reticulum (ER): pH 7.0-7.4[114-116], Golgi: pH 6.1-6.7[114-116], Silica Deposition Vesicle (SDV): pH < 7[117]). Recombinant tpSTK1 is only active in a narrow pH range, from 7-8, in Tris based buffers (Figure 3.13 C). In the phosphate/citrate buffer system, using the same amount of magnesium as in the kinase assay buffer (10 mM), no kinase activity is observed (data not shown). It is possible that the citrate in the buffer chelates the magnesium ions, making them unavailable for the kinase to use for its activity. To compensate for the loss of magnesium due to complexation with citrate, a 10 fold increase in the Mg^{2+} concentration was included in the kinase buffer and resulted in significant kinase activity at pH 7 and 8. Thus, recombinant tpSTK1 is active in the neutral to slightly basic pH range in various buffers. Since the ER is neutral and the Golgi and SDV are more acidic, one can speculate that if tpSTK1 is a membrane bound kinase, it is likely to reside in the ER.

3.4 Purification and chemical analysis of phosphorylated recombinant silaffins

3.4.1 RP-HPLC purification and phosphate release determination

To determine the amount of phosphate incorporated into the recombinant silaffins, the recombinant tpSTK1-phosphorylated proteins rSil3 and rSilC were purified via RP-HPLC on a C_{18} column and then chemically analyzed for phosphate content.

Phosphorylated rSil3 (rSil3-P) and rSilC (rSilC-P) exhibited significant downshifts in their RP-HPLC elution times consistent with a decrease in hydrophobicity due to phosphorylation (Figure 3.14).

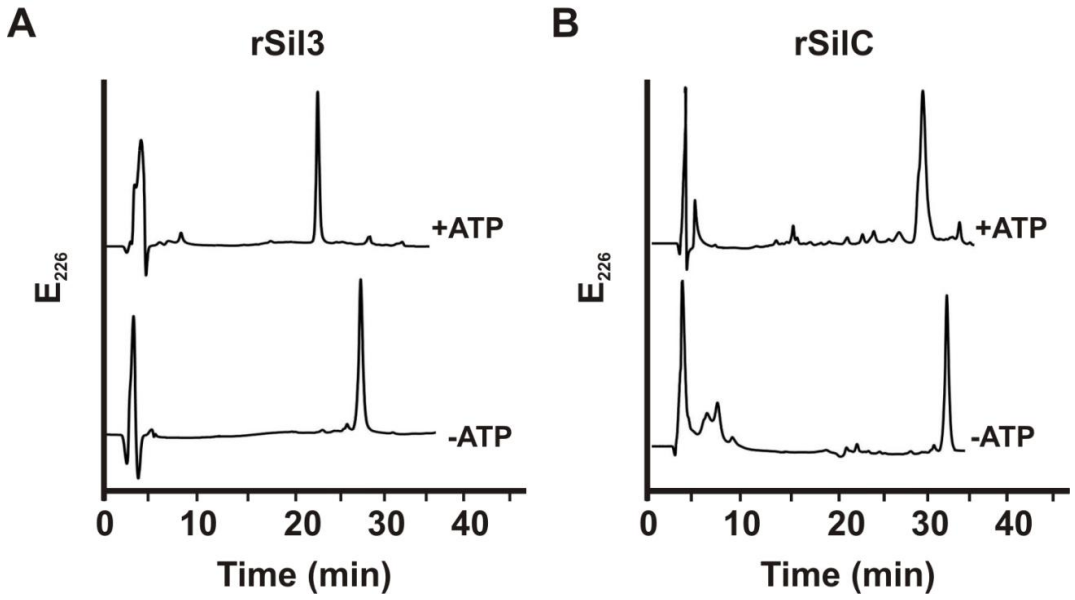


Figure 3.14. RP-HPLC of phosphorylated recombinant silaffin substrates[97]. Analysis after incubation of ***A***, rSil3 or ***B***, rSilC with tpSTK1 in the presence (+ATP) and absence of (-ATP).

To determine the number of phosphates incorporated on the recombinant silaffins by recombinant tpSTK1, a colorimetric phosphate release assay was performed. The dried protein residues are acid-hydrolyzed to release the phosphate from the protein backbone. A dye (malachite green in this case) then binds to the released phosphate under acidic conditions resulting in a color formation that can be detected photometrically and

the phosphate content calculated from a inorganic phosphate standard curve (Figure 3.15). As controls, ATP and glycerol phosphate were subjected to phosphate content chemical analysis and the number of phosphates detected for ATP and glycerol phosphate were 2.9 - 3.1 and 0.8 - 1.0 mol phosphate per mole protein, respectively (Table 3.3). These values are consistent with the known phosphate content of ATP (3 phosphates per molecule) and glycerol phosphate (1 phosphate per molecule), demonstrating the accuracy of the applied method. Chemical analysis of rSil3-P and rSilC-P detected 2.8 - 3.1 and 0.8 - 1.1 mol phosphate per mole of protein, respectively (Table 3.3). The number of phosphates incorporated on the recombinant silaffins is quite low given that native silaffins from *T. pseudonana* are estimated to contain 35 to 50 phosphate residues/molecule[61].

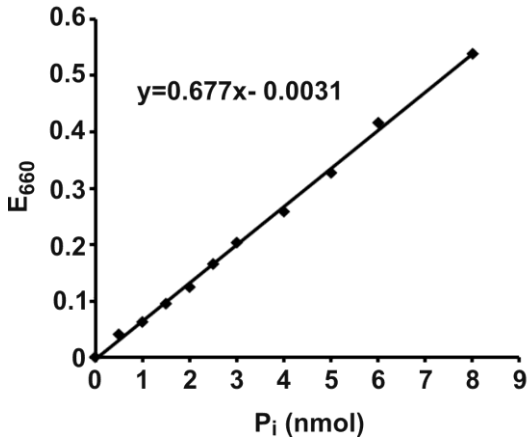


Figure 3.15. Quantification of inorganic phosphate using the malachite green method (see section 2.20.1)[97].

Table 3.3. Analysis of phosphate content in silaffins and reference compounds[97]. Phosphorylated silaffin rSil3 (rSil3-P) and phosphorylated silaffin rSilC (rSilC-P) were purified by RP-HPLC and quantified as described in section 2.20.1. ATP (contains 3 phosphate residues per molecule) and glycerophosphate (GlcP; contains 1 phosphate residue per molecule) were used as purchased. Inorganic phosphate obtained after hydrolysis was calculated according to the inorganic phosphate reference shown in Figure 3.15.

	molecule (nmol P_i)	**E_{660}**	**P_i calculated (nmol)**	**Pi:molecule**
rSil3-P	0.1	0.016	0.28	2.8
	0.75	0.152	2.30	3.1
	1.1	0.216	3.24	2.9
rSilC-P	0.17	0.012	0.14	0.8
	1.26	0.112	1.40	1.1
	1.85	0.178	1.79	1.0
GlcP	0.50	0.013	0.39	0.8
	2.00	0.113	2.01	1.0
	5.00	0.272	4.36	0.9
ATP	0.25	0.046	0.73	2.9
	0.50	0.102	1.56	3.1
	2.50	0.521	7.73	3.1

3.4.2 Phosphoamino acid analysis

Phosphoamino acid analysis of RP-HPLC-purified rSil3-P detected phosphoserine; consistent with the prediction that tpSTK1 is an S/T kinase (Figure 3.16). This finding is biologically relevant as phosphoserine is also present in the corresponding native silaffin, tpSil3, isolated from *T. pseudonana*. Native tpSil3 also contains phosphothreonine and *O*-sulfated serine (an amino acid artificially generated during acid hydrolysis of tpSil3), which are absent from recombinant tpSTK1-phosphorylated rSil3.

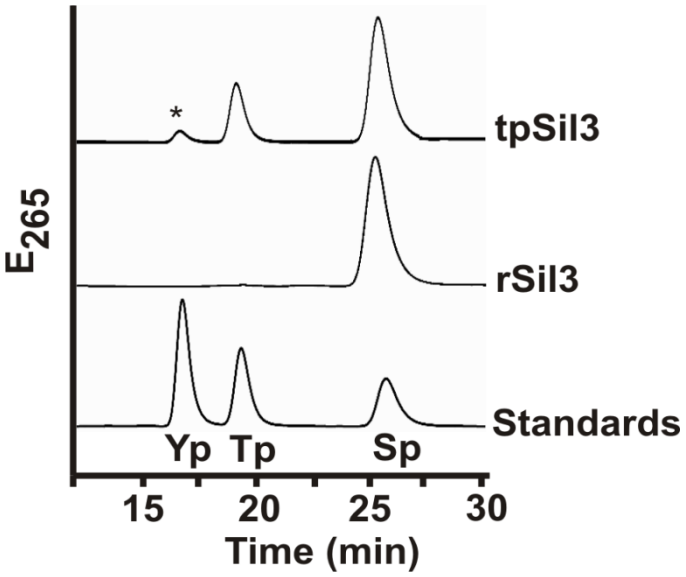


Figure 3.16. Chemical analysis of phosphorylated substrates[97]. Phosphoamino acid analysis by anion exchange HPLC of tpSTK1-phosphorylated rSil3 (rSil3-P) and native silaffin tpSil3. The peak labeled with an asterisk contains *O*-sulfated serine which was artificially generated during tpSil3 hydrolysis (N. Poulsen and N. Kröger, unpublished result). As a reference the chromatogram of a mixture of the standard amino acids phosphotyrosine (Yp), phosphothreonine (Tp) and phosphoserine (Sp) is shown.

3.5 tpSTK1 homolog, tpSTK2

3.5.1 Primary sequence alignments

The low number of incorporated phosphates and the lack of phosphothreonine in rSil3 treated with recombinant tpSTK1 suggest that other kinases may be involved in the phosphorylation of silaffins. A BLAST[107] search of the NCBI database revealed significant sequence alignments (e values < 2e-7) of tpSTK1 to more than 50 proven or predicted serine/threonine kinases from a broad range of organisms (including other biomineralizing organisms). The alignment to kinases from non-diatom organisms had

relatively high e-values (between 2e-10 to 8e-7), while three predicted diatom kinases, tp7188 (GenBank# ACI64649.1) from *T. pseudonana*, and pt45008 (GenBank# EEC49800.1) and pt32914 (GenBank# EEC50886.1) from *Phaeodactylum tricornutum* exhibited much lower e-values (3e-86, 1e-71, and 5e-44, respectively), and thus much higher sequence similarity (Figure 3.17). The three tpSTK1-like diatom kinases each contain a predicted kinase domain in the C-terminal part of the protein exhibiting 26.0 % - 28.8 % amino acid sequence conservation amongst each other and tpSTK1. In addition, the positions of 44.4% - 48.0 % of the amino acid residues within the kinase domains are conserved amongst at least three of the predicted diatom kinases and tpSTK1.

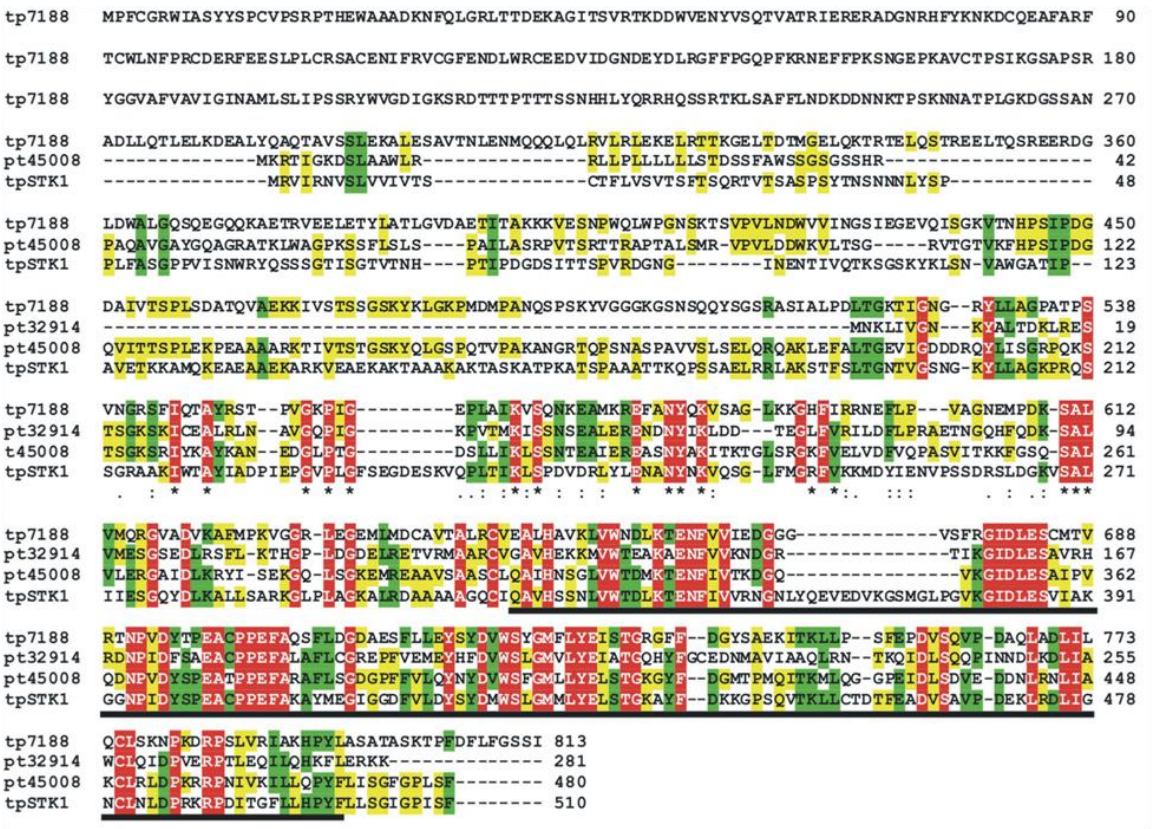


Figure 3.17. **Primary structures of tpSTK1 and related proteins[97].** Alignment of tpSTK1 with predicted protein tp7188 (GenBankTM accession number ACI64649.1) from *T. pseudonana* and predicted proteins pt45008 and pt32914 (GenBankTM accession numbers EEC49800.1 and EEC50886.1) from *P. tricornutum*. The alignment was performed using the ClustalW2 program[118]. Amino acids that are conserved among all (red), three (green), and two (yellow) sequences are highlighted. The predicted kinase domain of tpSTK1 is underlined.

3.5.2 Substrate specificity of recombinant tpSTK2-KD

The *T. pseudonana* tpSTK1 homolog tp7188, denoted in the following as tpSTK2, was investigated to determine its involvement, if any, in silaffin phosphorylation. Since the full length cDNA sequence of the *tpstk2* gene predicted in the *T. pseudonana* genome database, could not be confirmed by RACE-PCR and RT-PCR (Appendix J), the region of the gene encoding for the putative kinase domain of tpSTK2 (which was confirmed by

RACE-PCR and RT-PCR) was fused to a hexahistidine tag, expressed in *E. coli* and solubilized from inclusion bodies. Recombinant tpSTK2-KD (33 kDa, pI = 6.03, Appendix K) was purified to homogeneity by immobilized metal affinity chromatography under denaturing conditions (8 M urea) (Figure 3.18) then subjected to on-column refolding under the same conditions used to yield active recombinant tpSTK1 (see section 2.13.2 for details).

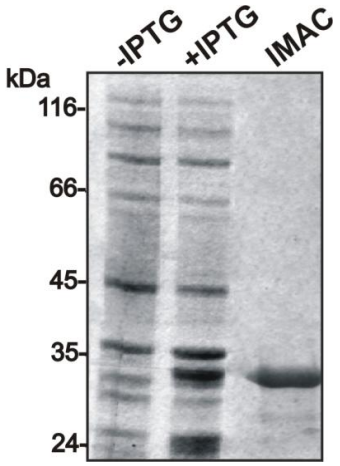


Figure 3.18. Expression and isolation of recombinant tpSTK2 Kinase Domain. Coomassie-stained SDS-PAG. Identical amounts of cells from an *E. coli* clone harboring a pET16b plasmid encoding recombinant tpSTK2-KD were loaded before (-) and after 3 hours of induction with IPTG (+). After solubilization of the inclusion bodies with urea, adsorption of the solubilized material to Ni-NTA agarose, and refolding on the column, the majority of recombinant tpSTK2-KD eluted with 200 mM imidazole (immobilized metal affinity chromatography = IMAC).

Recombinant tpSTK2-KD was tested for its activity with the recombinant silaffins, recombinant silacidin and the commercial proteins as substrates using the luminescence based kinase assays. Recombinant tpSTK2-KD only exhibited kinase activity on myelin basic protein (Figure 3.19, black bars). It is possible that recombinant

tpSTK2-KD phosphorylates the other substrates only after previous phosphorylation of the substrates with another kinase. For example, casein kinase I is active on recombinant phosphophoryn only after prior phosphorylation with casein kinase II[76]. To test whether tpSTK2-KD is active on phosphorylated proteins, the substrates were first incubated with tpSTK1 for one hour, and then tpSTK2 was added and the reaction mixtures were incubated for an additional hour. The phosphorylation of the substrates due to recombinant tpSTK2 activity was calculated by subtracting the tpSTK1 activity. Most of the substrates were not phosphorylated by recombinant tpSTK2-KD, even after prior phosphorylation with tpSTK1 (Figure 3.19, pink bars). However, the activity of recombinant tpSTK2-KD on phosphorylated myelin basic protein (MBP-P) increased by 42 %, suggesting a synergistic effect of tpSTK1 and tpSTK2 on myelin basic protein. Although tpSTK2-KD did not phosphorylate any of the recombinant silaffins or recombinant silacidin, these results show that tpSTK2-KD is a protein kinase.

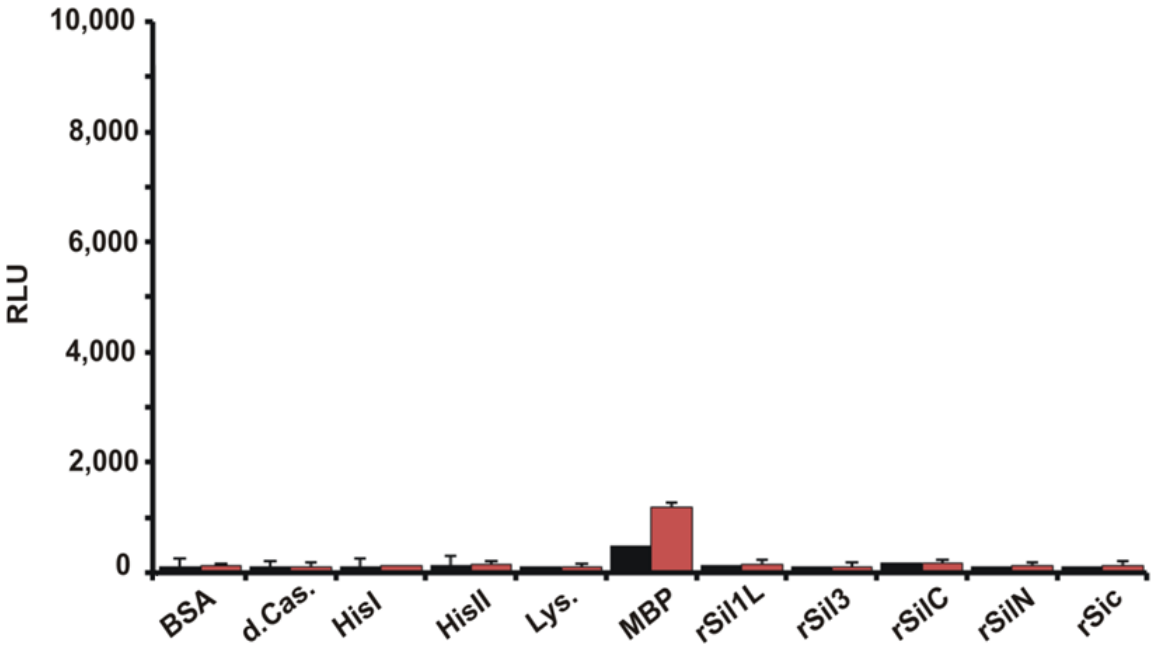


Figure 3.19. Phosphorylation activity of recombinant tpSTK2-KD with recombinant silaffins (rSilN, rSilC, rSil3, and rSil1L), recombinant silacidin (rSic), and commercial proteins as substrates. ***Black bars***, Luminescence-based kinase assays with 0.10 mg/mL substrate protein (black bars). ***Grey bars***, Substrates were phosphorylated by tpSTK1 first before tpSTK2-KD activity assays. TpSTK1 activity has been subtracted to yield only the kinase acitivty due to tpSTK2-KD. Each measurement was repeated at least three times. RLU, relative luminescence units. Lys- Lysozyme, His-Histone, MBP-Myelin Basic Protein, d. Cas- dephosphorylated casein, BSA- bovine serum albumin.

3.6 Intracellular location of native tpSTK1[97]

3.6.1 $TpSTK_{1\text{-}24}$: N-terminal ER targeting peptide or transmembrane domain?

As mentioned earlier, current sequence analysis tools cannot unequivocally distinguish between a transient targeting peptide for co-translational ER import and a permanent membrane anchor[110]. The ability of the first 24 amino acids of tpSTK1 to act as N-terminal ER signal peptide was evaluated by observing (using microscopy) the intracellular localization of green fluorescent protein with ($STK_{1\text{-}24}GFP_{DDEL}$) and without

(GFP_{DDEL}) the N-terminal tpSTK1 peptide in *T. pseudonana*. The tetrapeptide D-D-E-L, which is the ER retention signal from the *T. pseudonana* ER-resident molecular chaperone BiP (binding protein)[119], was added at the C-terminus of the GFP in both constructs to prevent its secretion into the medium by retaining it in the ER. The GFP_{DDEL} protein that was expressed without an N-terminal extension exhibited a cytosolic localization pattern whereas the GFP_{DDEL} fused to STK_{1-24} was located in a clamp-like structure that was associated with the chloroplast (Figure 3.20 top). A similar localization pattern was observed with GFP fused to the signal peptide of the ER resident protein, BiP ($BiP_{1-27}GFP$). As diatom chloroplasts are tightly surrounded by ER membranes[120], the chloroplast-associated clamp-like structures containing $STK_{1-24}GFP_{DDEL}$ and $BiP_{1-27}GFP$ may represent sub-compartments of the ER.

During maturation, the N-terminal signal peptides of proteins destined for the ER are cleaved off by endogenous signal peptidases, a characteristic of an N-terminal signal peptide that is not shared with transmembrane domains. Western blot analysis revealed that the apparent molecular mass of $STK_{1-24}GFP_{DDEL}$ in *T. pseudonana* was identical to that of the cytosolically expressed GFP_{DDEL} indicating that the N-terminal STK_{1-24} part was cleaved off (Figure 3.20 bottom). The same is true for the $BiP_{1-27}GFP$ fusion protein as the resulting protein, GFP, was smaller in apparent molecular mass than GFP carrying the DDEL tetrapeptide.

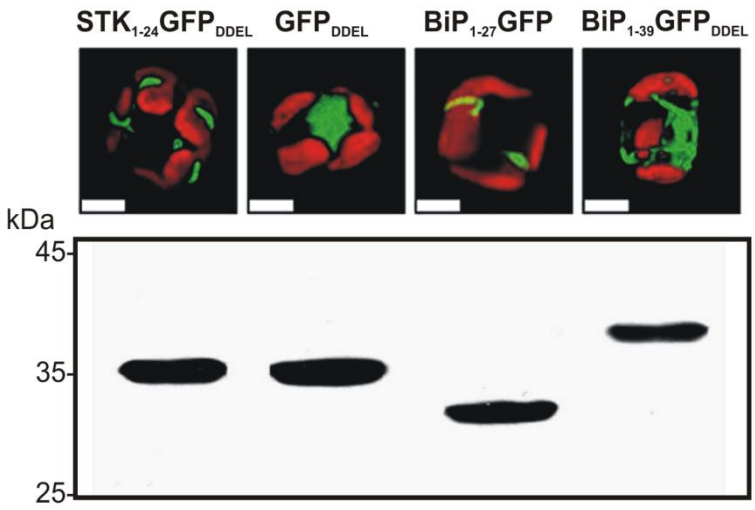


Figure 3.20. Intracellular processing and localization of GFP fusion proteins[97]. For names and composition of the GFP fusion proteins refer to section 3.22. ***Top***, Confocal fluorescence micrographs of individual *T. pseudonana* cells expressing the GFP fusion proteins indicated in the top panel. The red color is caused by autofluorescence of the chloroplasts. Scale bars = 2 µm. Confocal images taken by Nicole Poulsen. ***Bottom***, corresponding Western blot with anti-GFP antibodies (0.5 µg/mL). In each lane, whole cell lysates of *T. pseudonana* expressing the indicated GFP fusion protein were loaded. After removal of the N-terminal signal peptide from $BIP_{1–39}GFP_{DDEL}$, 12 amino acids derived from the mature BiP sequence remain attached to the N terminus of GFP.

A second BiP-GFP fusion protein contained amino acids 1-39 and the C-terminal ER-retention signal of BiP. Removal of the signal peptide during intracellular maturation was expected to generate a GFP fusion protein that contains 12 and 16 amino acids more than the proteins derived from $STK_{1-24}GFP_{DDEL}$ and $BiP_{1-27}GFP$, respectively. Indeed, Western blot analysis confirmed that the $BiP_{1-39}GFP_{DDEL}$ had the highest molecular mass of all GFP fusion proteins that were analyzed (Figure 3.20 bottom). Interestingly, the $BiP_{1-39}GFP_{DDEL}$ derived protein exhibited an ER-like localization pattern suggesting that the additional amino acids derived from mature BiP have a strong influence on the intracellular localization (Figure 3.20 top). Altogether, these data indicate that amino

acids 1-24 of tpSTK1 function as a cleavable signal peptide for import into the ER and not a permanent membrane anchor domain.

3.6.2 Localization of native tpSTK1

3.6.2.1 GFP tagged tpSTK1

The signal peptides of tpSTK1 and BiP targeted GFP to the chloroplast associated clamp-like structures (that may be ER sub-compartments), yet only a few additional amino acids influenced the localization pattern of the BiP protein. Therefore, additional targeting signals present in the full length proteins must be present to direct the proteins out of the ER sub-compartments and into other areas within the cell. To determine the intracellular location of native tpSTK1, the full length tpSTK1 gene fused to GFP fusion gene was constructed and introduced into *T. pseudonana* cells. For comparison, the full length ER BiP protein fused to GFP was also introduced into *T. pseudonana*. Although the cells expressed tpSTK1-GFP, as evident by the detection of an 85 kDa band on a Western blot probed with anti-GFP IgGs (Figure 3.21 A), no GFP-fluorescent transformants were obtained for tpSTK1-GFP. If tpSTK1 was located in the ER, the tpSTK1-GFP fusion protein would be expected to show a reticulate-like localization pattern that is observed with the BiP-GFP fusion protein (Figure 3.21 B). One possible reason for the lack of GFP fluorescent transformants may be improper or partial folding of the fusion protein. However, placing the tag on the N-terminal end of the protein may promote proper folding. Therefore, a fusion construct was generated that placed the GFP near the N-terminus of the presumed mature tpSTK1 (Figure 3.21 C). However, no GFP-transformants were obtained, although the fusion protein was expressed.

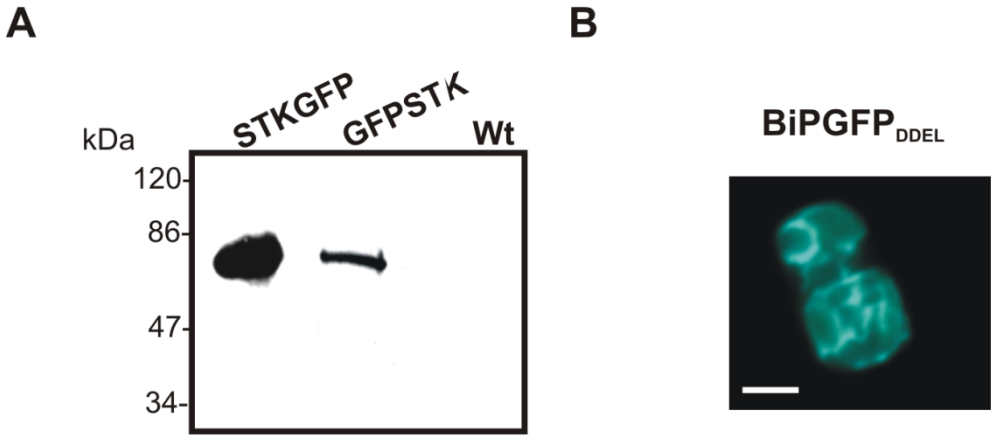


Figure 3.21. TpSTK1-GFP fusion constructs expressed in the diatom *T. pseudonana*. ***A***, Western blots with anti-GFP antibodies (0.5 µg/mL) of cells expressing tpSTK1 with a C-terminal GFP (STK-GFP) and an N-terminal GFP (GFP-STK1). Wild type cells (**Tpwt**) were included as a control. Equal aliquots of cells were loaded. ***B***, Epifluorescence micrograph of a *T. pseudonana* cell expressing the full-length BiP protein fused to a C-terminal GFP tag with the diatom ER retention tetrapeptide (DDEL). Scale bar = 2 µm.

3.6.2.2 Biochemical localization of native tpSTK1

As GFP tagging of tpSTK1 was unsuccessful, a biochemical approach was pursued for identifying the intracellular localization of native tpSTK1. *T. pseudonana* cells were gently lysed with glass beads and the lysate separated into three fractions by differential centrifugation: P1 (16,000*g* pellet), P2 (100,000*g* pellet), and S2 (100,000*g* supernatant). S2 is the cytosolic fraction, P1 contains non-lysed cells, cell walls, and ~80 % of the chloroplasts (estimation based on chlorophyll content), and P2 contains plastid-depleted cellular membranes. Western blot analysis with anti-tpSTK1 IgG revealed that tpSTK1 is predominantly present in the P2 fraction and virtually absent from the cytosol. TCA/DOC precipitation of proteins from the diatom medium (see section 2.26 for

experimental details) show that tpSTK1 is not secreted into the medium (Figure 3.22 Left). Given the intimate association of the ER with diatom plastids, it is likely that proteins within the chloroplast associated-clamplike structure (CA-CLS) would co-localize with diatom plastids. However, the localization of mature tpSTK1 with plastid-depleted cellular membranes suggests that tpSTK1 does not reside in the CA-CLS. The P2 membranes were fractionated by sucrose density gradient centrifugation (optimized to separate cellular membranes from each other) yielding fractions F1-F7 (Figure 3.22 right).

Figure 3.22. Intracellular location of native tpSTK1 in *T. pseudonana* wild type cells[97]. *Left*, Western blot probed with anti-tpSTK1 IgG (0.2 µg/mL) showing the distribution of native tpSTK1 in subcellular fractions and the culture medium. Equal aliquots of whole cell lysate (Cells), cell culture medium (sec.), and subcellular fractions (P1, P2, S2) were subjected to SDS-PAGE. ***Right,*** Schematic of sucrose density gradient. Fractions (F1-F7) were collected after centrifugation.

Each fraction was analyzed for the presence of tpSTK1, protein and chlorophyll content, and the activities of marker enzymes for ER (NADH: cytochrome c reductase = CCRase), mitochondria (cytochrome c oxidase = COx), Golgi apparatus (IDPase), and plasma membranes (vanadate-sensitive ATPase = vsATPase). According to the marker

enzyme assays fraction F4 contains the majority of the ER membranes (71 % of total CCRase activity), almost all of the Golgi membranes are present in fraction F5 (74 % of total IDPase activity), and combined fractions F6 and F7 contain the majority of mitochondria (79 % of total COx activity) and plasma membrane (96 % of total vsATPase activity) (Table 3.4).

Table 3.4. Enzyme activities and protein and chlorophyll content in subcellular fractions of *T. pseudonana*[97]. The listed numbers represent averages from three independent experiments. The number for the highest activity of a particular enzyme is shown in bold. IDPase, inositol diphosphatase; CCRase, cytochrome *c* reductase; COx, cytochrome *c* oxidase; vsATPase, vanadate-sensitive ATPase; Chl, chlorophyll; n.d., not determined.

Fraction	IDPase (nmol P_i/sec)	NADH:CCRase (nmol cyt c red/min)	COx (nmol cyt c ox/min)	vsATPase (nmol P_i/sec)	protein (mg)	chl. (mg)
S2	n.d.	n.d.	n.d.	n.d.	1.31±0.09	0.04±0.01
P1	n.d.	n.d.	n.d.	n.d.	0.79±0.06	2.45±0.13
P2	n.d.	n.d.	n.d.	n.d.	1.20±0.10	0.65±0.08
F1 + F2	0.04±0.01	0.07±0.03	0.04±0.02	0.01±0.00	0.41±0.03	0.11±0.03
F3	0.05±0.02	0.33±0.02	0.03±0.01	0.01±0.00	0.14±0.01	0.06±0.02
F4	0.34±0.03	**1.74±0.04**	0.03±0.00	0.01±0.01	0.21±0.01	0.03±0.02
F5	**1.44±0.02**	0.04±0.03	0.11±0.04	0.04±0.02	0.19±0.02	0.02±0.01
F6 + F7	0.06±0.00	0.08±0.02	**0.81±0.06**	**1.41±0.06**	0.34±0.01	0.40±0.04

Western blot analysis of the same gradient fractions using anti-tpSTK1 antibodies detected tpSTK1 mainly in fraction F4, and to a smaller extent in fraction F3, but not in any other fraction (Figure 3.23). The fact that the distribution of tpSTK1 within the sucrose gradient perfectly paralleled the distribution of ER membranes indicates that the kinase is likely associated with ER membranes *in vivo*.

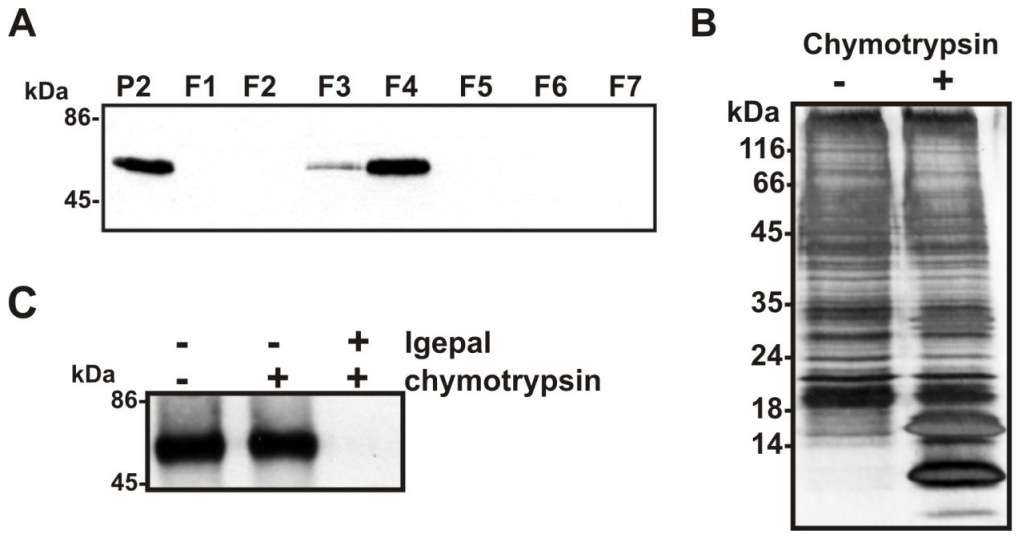


Figure 3.23. Distribution of tpSTK1 in subcellular membrane fractions[97]. ***A***, Western blot probed with anti-tpSTK1 IgG (0.2 µg/mL) of native tpSTK1 in membrane fractions from density sucrose gradient centrifugation. Equal aliquots of sucrose gradient fractions F1–F7 were loaded. ***B***, Chymotrypsin treatment of intact membranes**.** Silver-stained SDS PAG. F4 membranes (10 µg/mL total protein) were incubated for 1 hour at 4 ^{o}C in the absence (-) or presence (+) of 0.3 mg/mL chymotrypsin. ***C***, Orientation of native tpSTK1 in F4 membranes from *T. pseudonana*. Protease accessibility test. Equal amounts of membranes from fraction F4 were incubated in the absence (-) or presence (+) of the detergent Igepal and 1.0 mg/mL of the protease chymotrypsin.

3.6.3 Protease Accessibility Assays

As the silaffins are transported through the endomembrane system, they are not exposed to the cytosol. Therefore, tpSTK1 must reside within the lumen of the ER membranes if it is to interact with silaffins *in vivo*. To determine whether tpSTK1 is localized on the cytosolic or the luminal side of membranes, protease accessibility experiments were performed with the membranes isolated from fraction F4. Treatment of the membranes with chymotrypsin resulted in degradation of several major protein components (Figure 3.23), but did not disrupt the activity of the ER luminal enzyme

CCRase (Table 3.5). However, CCRase activity was completely disrupted when the membranes were detergent solubilized prior to chymotrypsin treatment. These results indicated that the surfaces of the isolated ER membranes were correctly oriented ("inside-in"). To test the accessibility of tpSTK1 to proteases, intact F4 and detergent-solubilized membranes were incubated with 1.0 μg of chymotrypsin for one hour on ice. If tpSTK1 is in the lumen of the ER membranes, then it should be protected from protease degradation in the same manner as was observed for CCRase. Western blot analysis of chymotrypsin-treated intact F4 membranes revealed that tpSTK1 is inaccessible to protease digestion. Degradation of tpSTK1 is only observed when the membranes are solubilized by the detergent Igepal, demonstrating that it is located in the membrane lumen (Figure 3.23).

Table 3.5. Protease accessibility of CCRase activity in F4 membranes[97]. The activity of this ER-located enzyme was measured after treatment of intact and detergent-solubilized (1 % w/v Igepal) F4 membranes with 1.0 mg/mL chymotrypsin. The listed numbers represent the averages from three independent experiments.

	CCRase (nmol cyt c_{red}/min)
intact membranes	1.81
intact membranes + chymotrypsin	1.75
solubilized membranes + chymotrypsin	0.01

3.6.4 Amount of native tpSTK1 in the ER

To estimate the amount of native tpSTK1 in the ER membranes, Western blot analysis comparing the intensity of various amounts of F4 membranes to known

concentrations of recombinant tpSTK1 was performed. The intensity of the signal of native tpSTK1 in the F4 fraction containing 10 µg of total protein is almost the same intensity as that observed from 0.05 µg of recombinant tpSTK1 (Figure 3.24). Based on this observation, it is estimated that 0.5 % of the total protein content in the ER fraction is native tpSTK1, making it a relatively abundant protein in the ER lumen.

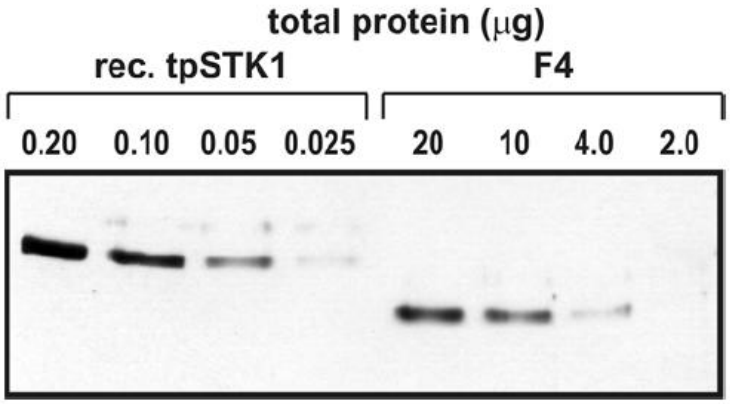


Figure 3.24. Western blot for estimation of native tpSTK1 content of F4 membranes[97]. Recombinant tpSTK1 was isolated by IMAC and gel extraction after SDS-PAGE. Protein concentrations were determined using the BCA assay[94]. The Western blot was probed with 0.2 µg/mL anti-tpSTK1 IgG.

3.7 TpSTK2 intracellular location

3.7.1 Location of native tpSTK2 in cellular membrane fractions

Using the established biochemical fractionation techniques, the intracellular location of tpSTK2 was explored. A polyclonal antiserum was produced against recombinant tpSTK2-KD and anti-tpSTK2 IgG molecules were obtained from the

antiserum by dual affinity chromatography on a recombinant tpSTK2-KD-loaded column followed by protein G-agarose. The anti-tpSTK2 IgGs detected a protein of apparent molecular mass of 90 kDa in a *T. pseudonana* whole cell lysate (Figure 3.25 A). Although the full length cDNA sequence of tpSTK2 is unknown, the gene model in the *T. pseudonana* encodes for a protein with a predicted molecular mass of 94 kDa. Therefore, the 90 kDa protein band detected in the diatom cells is presumed to be native tpSTK2. However, other bands estimated at 80, 50, 44, 40 kDa are detected by anti-tpSTK2 that could be degradation products of tpSTK2. Western blot analysis using anti-tpSTK2 IgG revealed that tpSTK2 is mainly present in the cytosolic fraction (S2) (Figure 3.25 B top), but a significant amount is also present in membrane fraction (P2). TpSTK2 is not secreted into the medium. Western blot analysis with anti-tpSTK1 IgG on the same fractions revealed that the localization of tpSTK1 with cellular membranes was consistent with previous results (Figure 3.25 B bottom).

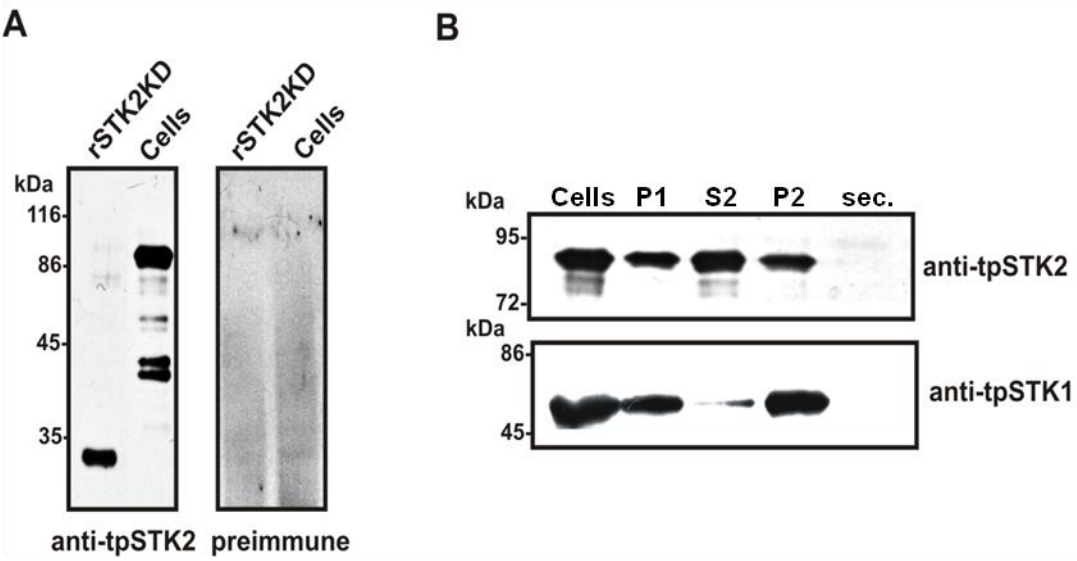


Figure 3.25. Intracellular location of native tpSTK2 and tpSTK1 in *T. pseudonana* wild type cells. ***A,*** Western blots probed with anti-tpSTK2 IgG (0.3 µg/mL) and preimmune IgG (0.3 µg/mL) that were isolated from the same rabbit. Lanes labeled "rSTK2"contained 200 ng of purified recombinant tpSTK2-KD; lanes labeled "Cells" contained lysates from 1.3×10^6 *T. pseudonana* cells. ***B,*** Western blots probed with anti-tpSTK2 (0.3 µg/mL) (***Top***) or anti-tpSTK1 IgG (0.2 µg/mL) (*Bottom*) showing the distribution of native tpSTK1 and tpSTK2 (the 86 kDa band) in subcellular fractions and the culture medium. Equal aliquots of whole cell lysate (cells), cell culture medium (sec.), and subcellular fractions (P1, P2, S2) were subjected to SDS-PAGE.

Since the majority of tpSTK2 was present in the cytosolic fraction (S2), marker enzyme assays were performed with the cytosolic fraction to determine whether any membrane bound proteins (i.e. the marker enzymes) are present. The majority of the marker enzyme activity observed in S2 is from IDPase, the Golgi marker protein. There is some ATPase activity (plasma membrane) but virtually no activity from CCRase (ER marker) or COx (mitochondrial marker). The 2.6-fold more IDPase activity in the cytosolic S2 fraction than in the Golgi fraction F5 (Table 3.6) indicates that the Golgi membranes are not preserved during membrane preparation.

Table 3.6. Enzyme activities in *T. pseudonana* cytosolic fraction S2 and membrane fraction F5. The listed numbers represent averages from three independent experiments. IDPase, inositol diphosphatase; CCRase, cytochrome *c* reductase; COx, cytochrome *c* oxidase; vsATPase, vanadate-sensitive ATPase.

	S2	**F5**
IDPase (nmol P_i/sec)	3.48 ± 0.11	1.34 ± 0.06
NADH:CCRase (nmol cyt c red/min)	0.13 ± 0.11	0.03 ± 0.11
COx (nmol cyt c ox/min)	0.07 ± 0.11	0.09 ± 0.11
vsATPase (nmol P_i/sec)	0.43 ± 0.11	0.12 ± 0.11

Despite the small amount of tpSTK2 in the P2 membranes (relative to the cytosolic fraction), the membranes were fractionated by sucrose density gradient centrifugation and each fraction was analyzed by Western blot for the presence of tpSTK1 and tpSTK2. As observed previously, tpSTK1 is present mainly in F4 and to a smaller extent in F3 (Figure 3.26 bottom), co-localizing with the ER membranes. TpSTK2, however, is mainly present in F5, the Golgi containing membrane fractions, and to a lesser extent in F6 which contains mitochondria and plasma membranes but virtually no Golgi membranes (Figure 3.26 top). Although some tpSTK2 is present in the F6 fraction, the majority of tpSTK2 (possibly ~90 %) is distributed to the Golgi membrane fraction F5 and to the cytosolic fraction S2 which closely follows the distribution of the Golgi enzyme IDPase. Therefore, it is likely that tpSTK2 is a Golgi-localized kinase.

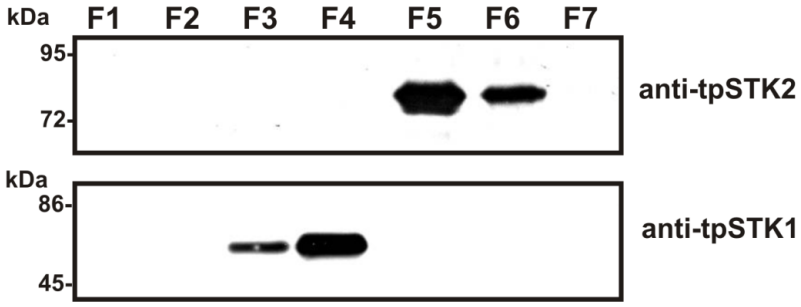


Figure 3.26. Distribution of tpSTK2 and tpSTK1 in subcellular membrane fractions. Western blot probed with anti-tpSTK2 (0.3 µg/mL) (***Top***) or anti-tpSTK1 IgG (0.2 µg/mL) (***Bottom***) of native tpSTK1 and tpSTK2 in membrane fractions from density sucrose gradient centrifugation. Equal aliquots of sucrose gradient fractions F1–F7 were loaded.

3.7.2 Protease Accessibility Assays

Protease accessibility assays were performed with the isolated membranes in F5 to determine if tpSTK2 is in the lumen of the Golgi and thus protected from protease degradation. Treatment of the membranes with chymotrypsin resulted in ~20 % loss of the total IDPase activity, revealing that some IDPase is exposed to the cytosol. However, the remaining 80 % of the total IDPase activity is lost only after the membranes are solubilized with detergent, indicating that there are a significant amount of intact Golgi membranes in this fraction (Table 3.7).

Table 3.7. Protease accessibility of IDPase activity in F5 membranes. The activity of this Golgi-located enzyme was measured after treatment of intact and detergent-solubilized (1 % w/w Igepal) F5 membranes with 1.0 mg/mL chymotrypsin. The listed numbers represent the averages from three independent experiments.

	IDPase (nmol P_i/sec)
intact membranes	1.49
intact membranes +chymotrypsin	1.21
solubilized membranes + chymotrypsin	0.02

Western blot analysis of the intact F5 membranes revealed that nearly 80 - 90 % of tpSTK2 was degraded by chymotrypsin (1.0 µg/µL) in one hour (Figure 3.27). However, a small amount of tpSTK2 is protected by from protease treatment that is not degraded even after a longer incubation (2 hours) with the protease. However, with the poor isolation of intact membranes and the presence of some membrane bound tpSTK2, the intracellular localization of tpSTK2 is uncertain.

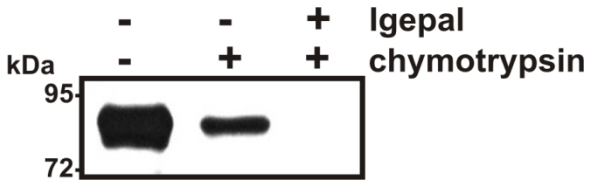


Figure 3.27. Orientation of native tpSTK2 in F5 membranes from *T. pseudonana*[97]. Protease accessibility test. Equal amounts of membranes from fraction F5 were incubated in the absence (-) or presence (+) of the detergent Igepal and 1.0 mg/mL of the protease chymotrypsin.

3.8 Kinase activity in ER and Golgi Membranes

3.8.1 Total kinase activity

To investigate the phosphorylation of silaffins and silacidin by kinases associated with the ER and Golgi apparatus, kinase assays were performed with membranes from sucrose gradient fractions F5 (contains Golgi membranes) and F4 (contains ER membranes) using recombinant silaffin substrates (rSil1L, and rSil3) and the recombinant silacidin. Kinase activity with all substrates was very low with intact membranes (Figure 3.28, black and grey bars) and at least 10-times higher with detergent-solubilized membranes (Figure 3.28, yellow and blue bars), indicating that the kinases acting on these substrates are luminal kinases. The kinase activity observed with substrate rSil3 and rSil1L was substantially higher with solubilized membranes from fraction F4 (contains ER membranes) than from fraction F5 (contains Golgi membranes), whereas the extent of rSic phosphorylation was almost identical with solubilized membranes from both fractions. Dephosphorylated casein was also tested as a substrate and phosphorylation of this substrate was observed from kinases in both ER and Golgi membrane fractions. This is consistent with previous reports that have shown that casein-like kinase activities are associated with ER and Golgi membranes from mammalian cells[74-76].

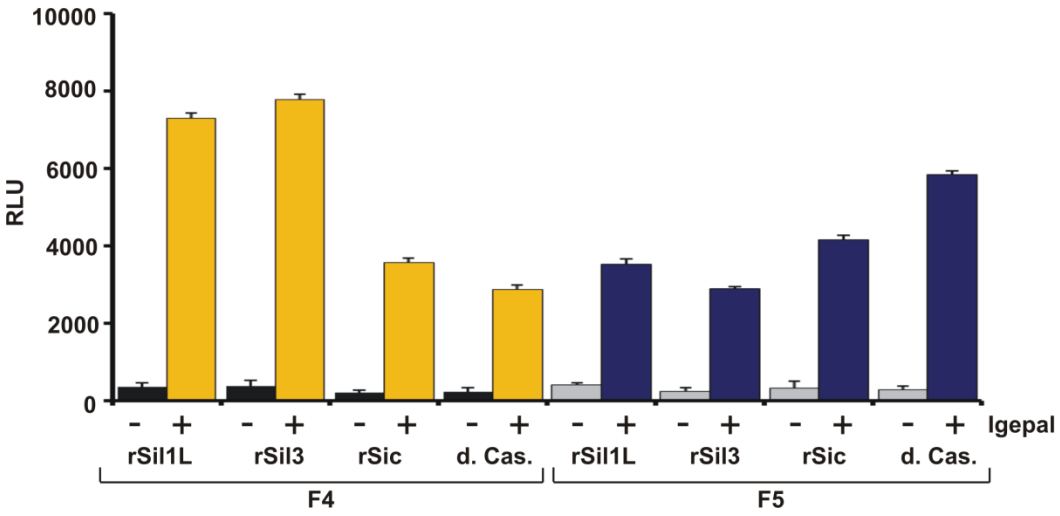


Figure 3.28. Phosphorylation of silaffins rSil3 and rSil1L, silacidin rSic, and dephosphorylated casein (Cas.) in subcellular fractions of *T. pseudonana* containing ER (F4) and Golgi (F5) membranes. For each assay equal amounts of membranes (10 µg of total protein) were used. The concentration of each substrate protein was 0.1 mg/mL . Kinase activities of intact membranes (-) and membranes solubilized with 1.0 % (v/v) Igepal (+) are shown. Each measurement was repeated at least three times and the standard deviation is indicated by ᴛ. RLU, relative luminescence units.

3.8.2 Kinase activity attributed to native tpSTK1

To determine the contribution of native tpSTK1 to the observed phosphorylation activities, kinase assays were performed in the presence of anti-tpSTK1 IgG. Before performing antibody inhibition assays with the F4 and F5 membrane fractions it was necessary to first determine the amount antibody required to completely inhibit tpSTK1 activity. To obtain this value, the activity of 100 µg/mL recombinant tpSTK1 on rSil3 was evaluated in the presence of increasing anti-tpSTK1 concentrations. At antibody concentrations ≥ 100 µg/mL, recombinant tpSTK1 (100 µg/mL) is virtually inactive (Figure 3.29 A). When the tpSTK1: IgG ratio remains constant (1:2 by mass), recombinant tpSTK1 is inactive at all protein concentrations indicating that the kinase-

antibody complexes are stable even at 10-fold dilution. Given that there is about 10 μg/mL of native tpSTK1 in the F4 fraction, incubating F4 and F5 membranes with 100 μg/mL of anti-tpSTK1 IgG (20:1 ratio by mass) should be more than sufficient to inhibit native tpSTK1 activity.

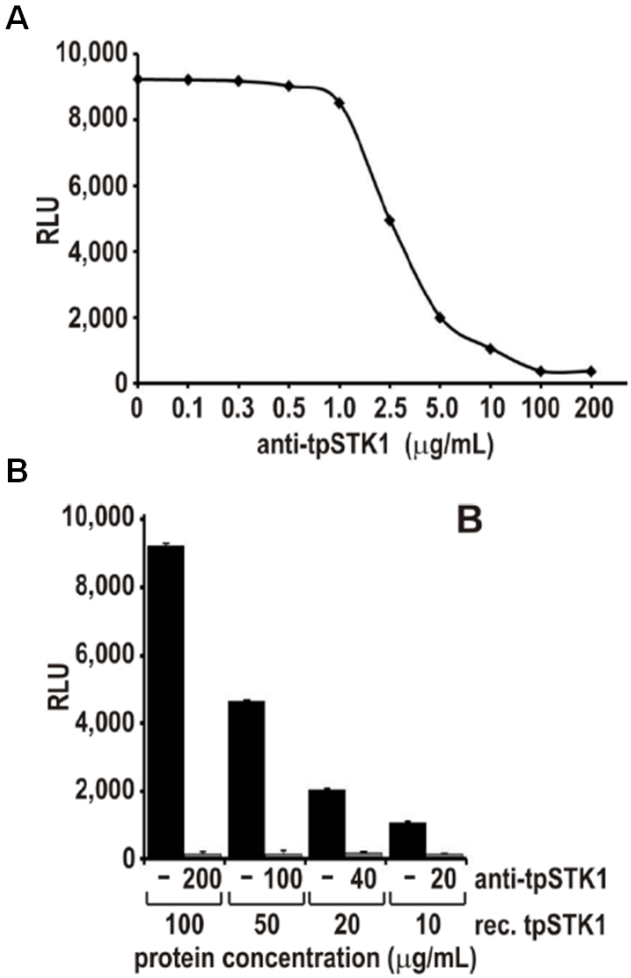


Figure 3.29. Inhibition of recombinant tpSTK1 by anti-tpSTK1 IgG *in vitro*[97]. Recombinant silaffin rSil3 was used as substrate protein. ***A***, Kinase activity of recombinant tpSTK1 (100 μg/mL) in the presence of increasing concentrations of anti-tpSTK1 IgG. Virtually complete inhibition of kinase activity (<5 % of maximum kinase activity) can be achieved at an antibody concentration of ≥100 μg/mL. ***B***, Kinase activity of recombinant tpSTK1 (rSTK1) in the presence of anti-tpSTK1 IgG at a constant tpSTK1:IgG ratio (1:2 by mass), but decreasing overall protein concentration. The kinase remained virtually inactive at all protein concentrations indicating that the tpSTK1:IgG complexes are stable even at high dilution.

In the presence of anti-tpSTK1 IgGs, the kinase activities of F4 and F5 membranes with substrates rSic, rSilN, and dephosphorylated casein remained virtually unchanged (Figure 3.30). With substrates rSil3, rSilC and rSil1L, the kinase activity of F5 membranes did not decrease in the presence of the anti-tpSTK1 IgGs, while the kinase activity of F4 membranes decreased by 24 %, 21 %, and 29 %, respectively. Using the same concentration of preimmune IgGs instead of anti-tpSTK1 IgGs did not inhibit the activities of the kinases present in solubilized membrane preparations from either fraction. These data indicated that native tpSTK1 was responsible for about 20 % - 30 % of the silaffin kinase activity in the ER lumen and was unable to phosphorylate acidic proteins rSilN, rSic, and dephosphorylated casein. The ability of native tpSTK1 to phosphorylate basic silaffin substrates and not acidic substrates (including rSic and dephosphorylated casein) is in perfect agreement with the phosphorylation activity of recombinant tpSTK1.

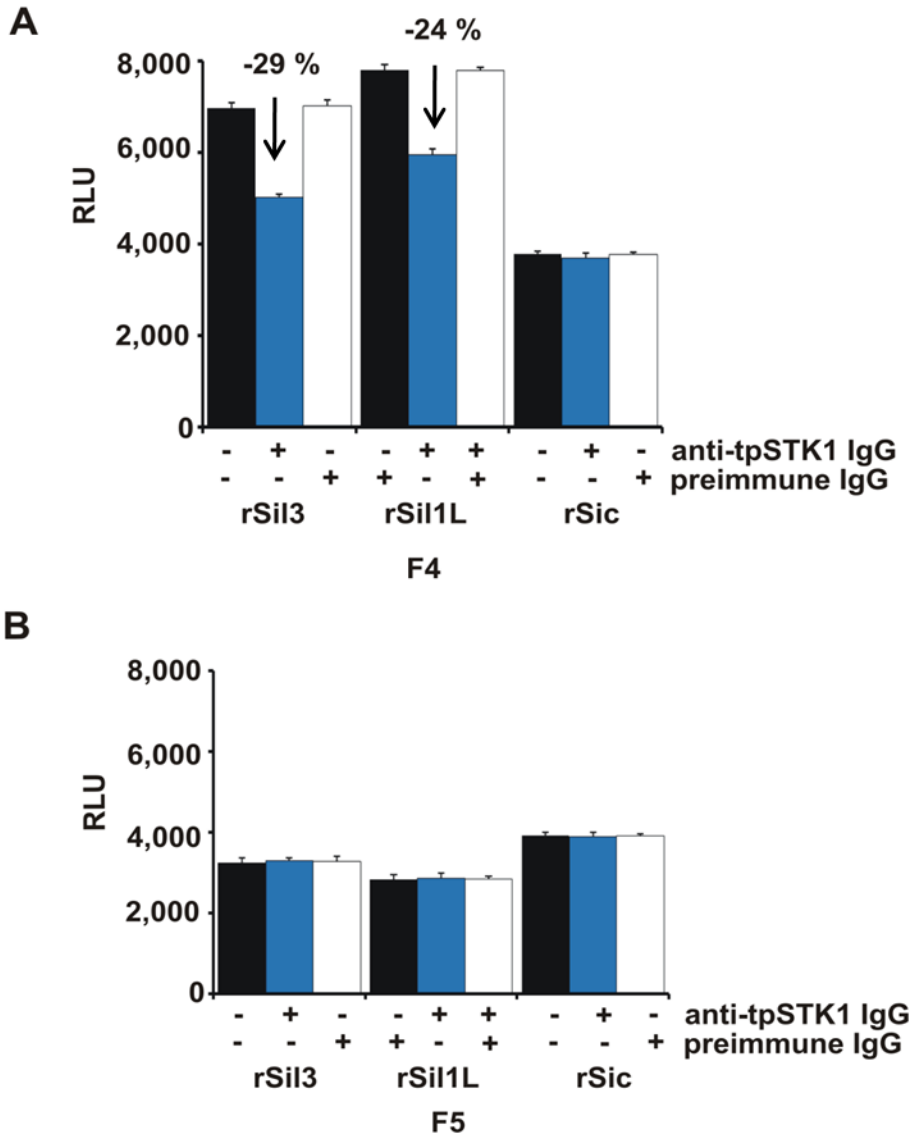


Figure 3.30. Inhibition of tpSTK1 activity in subcellular membrane fractions of *T. pseudonana*[97]. Phosphorylation of silaffins (rSil3 and rSil1L) and silacidin (rSic) in detergent-solubilized membranes of ER-containing fraction F4 (***A***) and Golgi-containing fraction F5 (***B***) in the presence (+) or absence (-) of equal concentrations of anti-tpSTK1 IgG or preimmune IgG. Each measurement was repeated at least three times. RLU, relative luminescence units.

3.8.3 Native tpSTK2 kinase activity

Antibody-based inhibition assays (as described for tpSTK1) were used to determine the contribution of native tpSTK2 to the observed phosphorylation activities; kinase assays were performed in the presence of anti-tpSTK2 IgG. The same amount (100 µg/mL) of antibody used to completely inhibit the activity of recombinant tpSTK1 *in vitro* was used in these assays. In the presence of the antibodies the kinase activities of F4 and F5 membranes remained virtually unchanged with all of the recombinant silaffin and silacidin substrates (rSilN, rSil3, rSilC, rSil1L and rSic) tested (Figure 3.31). The trivial amount of inhibition observed in the ER fraction with rSil1L as the substrate is within error and not attributed to tpSTK2 as tpSTK2 is not present in that fraction. Using the same concentration of preimmune IgG instead of anti-tpSTK2 IgG also did not inhibit the activities of the kinases present in solubilized membrane preparations from either fraction. These data indicate that native tpSTK2 is not a silaffin- (or a silacidin-) phosphorylating kinase, despite the similarities in its primary sequence to silaffin kinase tpSTK1.

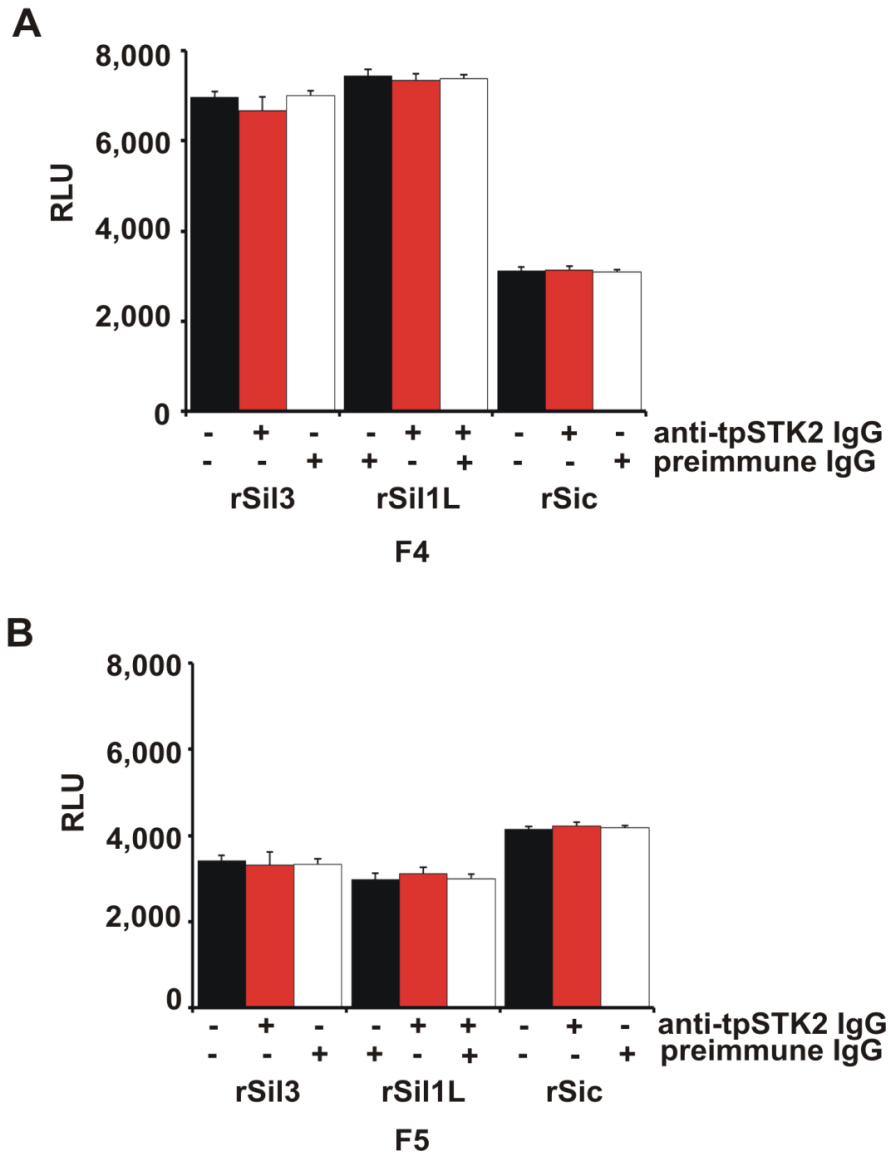


Figure 3.31. Inhibition of tpSTK2 activity in subcellular membrane fractions of *T. pseudonana*. Phosphorylation of silaffins (rSil3 and rSil1L) and silacidin (rSic) in detergent-solubilized membranes of ER-containing fraction F4 (***A***) and Golgi-containing fraction F5 (***B***) in the presence (+) or absence (-) of equal concentrations of anti-tpSTK2 IgG or preimmune IgG. Each measurement was repeated at least three times. RLU, relative luminescence units.

Antibody-based inhibition experiments using recombinant tpSTK2-KD and myelin basic protein as the substrate revealed that 100 µg/mL of anti-tpSTK2 IgG inhibited kinase activity by 61 % (Figure 3.32). As the same concentration of anti-tpSTK2 did not exert any inhibitory effect on phosphorylation of recombinant silaffin and silacidin by F5 membranes, it can be concluded that native tpSTK2 is incapable of phosphorylating these substrates.

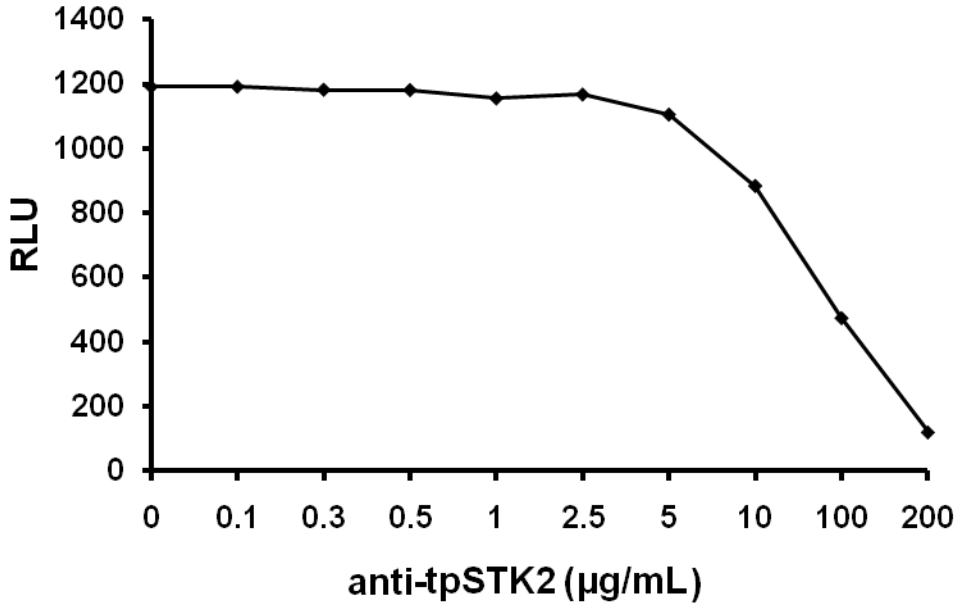


Figure 3.32. Inhibition of recombinant tpSTK2-KD by anti-tpSTK2 IgG *in vitro*. Myelin basic protein was used as substrate protein. Kinase activity of recombinant tpSTK2-KD (100 µg/mL) in the presence of increasing concentrations of anti-tpSTK1 IgG. Inhibition of kinase activity (61 % of maximum kinase activity) can be achieved at an antibody concentration of 100 µg/mL, the concentration used for assessing the native tpSTK2 kinase activity in the Golgi fraction.

CHAPTER 4: DISCUSSION

4.1 The Discovery of a Biomineralization kinase

Given the importance of phosphoproteins in biomineralization, the kinases that catalyze their phosphorylation are equally important, yet not a single biomineralization kinase had been identified from any organism (including humans!) until now. Of the two putative kinases, tpSTK1 and tpAK1, identified from a proteomics study on silica formation in the diatom *T. pseudonana*[65], only tpSTK1 emerged as a kinase capable of phosphorylating silaffin biosilicification proteins *in vitro*. TpSTK1 is a serine/threonine kinase that contains an N-terminal signal peptide for co-translational import into the endoplasmic reticulum (ER) where it remains in the lumen of this compartment (see Figures 3.20 and 3.22-3.24). Comparable to other abundant ER proteins such as protein disulfide isomerase (~ 0.4 % of total protein[121]) and BiP (estimates in the range of 1 % to 5 % of total protein[122]), tpSTK1 is a relatively abundant ER protein amounting to ~ 0.5 % of the total protein (see Figure 3.24). *In vitro* kinase assays revealed that tpSTK1 only catalyzed the phosphorylation of basic recombinant silaffin substrates (i.e. rSilC, rSil3, rSil1L, and R5) but not the acidic recombinant substrates, silaffin rSilN and silacidin rSic (see Figures 3.9-3.11). Although tpSTK1 phosphorylated the peptide substrate R5, the peptide had to be in 10-fold excess for it to be phosphorylated by recombinant tpSTK1, which indicates that tpSTK1 prefers proteins over peptides (see Figure 3.10). Recombinant tpSTK1 phosphorylation activity requires ATP, not GTP, as a co-substrate and prefers magnesium as its divalent metal cation co-factor, typical of most kinases[68]

(see Figure 3.13). Optimal kinase activity of recombinant tpSTK1 was observed in the pH range of 7-8, which is consistent with native tpSTK1 operating in the ER (which has a pH in the range of 7.0-7.4) (see Figure 3.13). Given the location and abundance of tpSTK1 within the ER and its phosphorylation of silaffin substrates *in vitro*, it is likely that tpSTK1 phosphorylates silaffins *in vivo*; therefore, tpSTK1 is regarded to be a silaffin kinase. TpSTK1 is the first structurally characterized kinase likely involved in biomineralization; and only the third kinase characterized that operates in the secretory pathway.

4.2 On the ER residency of tpSTK1

The exact location of tpSTK1 within the ER is unknown since expression of GFP-tagged tpSTK1, *in vivo*, has failed. The signal peptide of tpSTK1 targeted GFP to a chloroplast associated clamp-like structure (CLS), as did the signal peptide of a known ER resident protein, BiP. A structure similar to CLS, termed "blob"-like structure (BLS) located at the periphery of plastids, was observed by Kroth and co-workers in the diatom *Phaeodactylum tricornutum*[123-124]. BLS is an accumulation site for GFP-tagged nuclear-encoded plastid proteins, which contain an N-terminal signal peptide for ER import but lack a functional plastid import sequence (import of nuclear encoded proteins into diatom plastids proceeds through the ER, requiring a bipartite targeting sequence[120,125-126]). In diatoms, chloroplasts are tightly surrounded by the rough ER (termed CER)[119]. Therefore, CLS and BLS may represent subcompartments where ER-imported proteins transit before reaching their final destination. As the majority of native tpSTK1 was present in plastid-

depleted membranes, it is likely that mature tpSTK1 may not reside exclusively in CLS (, see Figure 3.22), but mainly in different regions of the ER. As the intracellular location of GFP-tagged tpSTK1 could not be confirmed with microscopy, immunolocalization techniques such as immunofluorescence microscopy and immunogold electron microscopy could be employed to elucidate the precise location of tpSTK1 within the ER.

TpSTK1 does not contain a C-terminal ER retrieval signal (C-terminal DDEL in diatoms, (H/K)DEL in other organisms), leaving the mechanism of tpSTK1 retention in the ER speculative. One possibility is that tpSTK1 contains an unknown ER retention motif. The ER-localized proteins s-cyclophilin (found in most eukaryotes) and a protein disulfide isomerase from a family of non-classical PDI-related proteins (PDI-D) from *Dictyostelium discoideum* both lack common ER retrieval signals[127-128]. Instead, s-cyclophilin contains a unique C-terminal extension (VEKPFAIAKE) and the *D. discoideum* PDI-D contains a 57 amino acid C-terminal domain (termed the D domain, hence the name PDI-D) for retention in the ER[127-128]. Such C-terminal ER retention motifs are not present in tpSTK1. Expression and localization of GFP-tagged truncations of tpSTK1 is required to identify the peptide segment of tpSTK1 responsible for ER retention..

Another possibility for retention of tpSTK1 in the ER may be that tpSTK1 does not contain an ER retention signal at all, but rather binds to other ER resident proteins (e.g. BiP and GRP94) that carry an ER retrieval signal. The retention of proteins in the ER by this mechanism has been observed in the thiol-mediated ER retention of Ero1α, a protein involved in protein disulfide isomerase oxidation[129]. The ER resident chaperone protein Erp44 contains the retrieval signal RDEL and forms reversible disulfide bonds

with Ero1α to prevent the secretion of Ero1α[129]. As Ero1α is an important protein in regulating oxidative protein folding (which occurs in the ER), the retention of this protein in the ER is critical for its function. If the retention of tpSTK1 in the ER is achieved by interactions with another ER resident protein, then it may be possible to capture these interactions and identify the tpSTK1-ER retaining protein using "pull down" methods and mass spectrometry. Using anti-tpSTK1 IgGs, immunoprecipitation can be employed to "pull down" tpSTK1and any bound proteins and the protein complexes analyzed by mass spectrometry. In the event that the tpSTK1-ER protein interaction is non-covalent or otherwise unstable, chemical crosslinking agents can be used to stabilize the protein-protein interaction due the presence of two or more reactive ends that chemically attach to specific functional groups on the proteins (i.e. primary amine groups, sulfhydryls). In addition to finding the tpSTK1-ER retaining protein (if such a protein exists), the combination of these methods would also lead to identification of the *in vivo* substrates of tpSTK1, which should include the *T. pseudonana* silaffins tpSil3 and tpSil1L.

4.3 Other Diatom Biomineralization Kinases

TpSTK1 is likely not the only kinase that phosphorylates silaffins *in vivo*. In the antibody based inhibition assays, native tpSTK1 only accounted for 25 % - 30 % of the silaffin kinase activity in the ER membranes, and there seem to be silaffin kinases present in the Golgi membranes (see Figure 3.29). Furthermore, recombinant tpSTK1 incorporated only 1-3 phosphates per silaffin molecule whereas native silaffins are estimated to contain 35 - 50 phosphate residues per molecule; and recombinant tpSTK1

specifically phosphorylated serine residues on rSil3 yet native tpSil3 also contains phosphothreonine in addition to phosphoserine (see Table 3.3 and Figure 3.16). To identify tpSTK1-related kinases, a BLAST[107] search using the primary sequence of tpSTK1 against three diatom genome databases was performed. The closest sequence homologues of tpSTK1 are the three predicted kinases, pt45008 (GenBankTM accession number EEC49800.1) and pt32914 (GenBankTM accession number EEC50886.1) from *P. tricornutum*, and tp7188 (GenBankTM accession number ACI64649.1) from *T. pseudonana.* Preliminary analysis of emerging genome sequencing data from the diatom *Fragilariopsis cylindrus* revealed five predicted kinases that exhibit higher sequence similarity to tpSTK1 and the tpSTK1-like diatom kinases (e-values: 8e-45 to 5e-70) than to any other serine/threonine kinases from diatoms or non-diatom organisms. These data suggested that tpSTK1 and its diatom homologs might represent a novel family of kinases involved in unconventional cellular pathways like silica biomineralization.

The only tpSTK1-like kinase from the diatom *T. pseudonana* is putative kinase tp7188, herein referred to as tpSTK2. Given its high primary sequence homology to tpSTK1 (~ 40 % identical), this kinase may also be involved in the phosphorylation of silaffins[97]. However, recombinant tpSTK2-KD was unable to phosphorylate recombinant silaffins or recombinant silacidin *in vitro* (see Figure 3.19). In the antibody-based inhibition kinases assays, native tpSTK2 apparently did not account for any of the phosphorylation activity on recombinant silaffins or recombinant silacidin from the ER and Golgi membrane fractions (see Figure 3.31). Taken together, these data suggested that tpSTK2 is not functionally related to tpSTK1. It should be noted, however, that

tpSTK2 cannot be completely ruled out as a diatom biomineralization kinase for the following reasons:

1.) The recombinant version of tpSTK2 that was expressed here (tpSTK2-KD) covered only of the kinase domain. The full length tpSTK2 may contain additional substrate recognition sequences that are not present in recombinant tpSTK2-KD.

2.) The anti-tpSTK2-KD antibodies used for the inhibition assay with Golgi containing membranes may not efficiently inactivate native tpSTK2.

3.) Native tpSTK2 may only act on substrates after they have been phosphorylated by other kinases or require the presence of other posttranslational modifications (e.g., alkylation of lysine residues, glycan moieties). The requirement for existing phosphorylation appears unlikely, because tpSTK1 was unable to phosphorylate recombinant silaffins that have been phosphorylated bytpSTK1.

4) TpSTK2 may phosphorylate hyrdoxyamino acids (e.g. hydroxyproline, dihydroxyproline) that are not present in the *E. coli* expressed recombinant silaffins. This presumption stems from the identification of phosphorylated hydroxyproline in the silaffin natSil2 from the diatom *Cylindrotheca fusiformis*[60]; similar phosphorylations on rare amino acids may also exist in *T. pseudonana* silaffins.

In addition to the tpSTK1-like kinases identified in the genomes of other diatoms, there are other putative kinases that may phosphorylate silaffins and/or silacidins. In a whole genome expression profiling study of *T. pseudonana* under different nutrient-limited conditions by Mock and coworkers, more than 10 genes encoding for putative kinases (not related to tpSTK1) showed expression patterns similar to that of tpSil3[91]. Ongoing research by Hildebrand and coworkers to identify proteins involved in silica

formation using whole-genome microarray analysis on synchronized *T. pseudonana* cells (unpublished data) revealed another six putative kinases, different from the 10 putative kinases in the Mock study, that may phosphorylate silaffins and/or silacidins. GFP-tagged variants of these candidate kinases could be expressed in *T. pseudonana* to determine their intracellular location via fluorescence microscopy. As evident by the kinase activity observed from ER and Golgi membrane fractions on recombinant silaffins and recombinant silacidin, and the presence of *O*-linked glycosylations on native silaffins, diatom biomineral forming proteins enter the secretory pathway via an ER targeting signal sequence and become modified within the endomembrane system (i.e. ER and Golgi) before reaching the silica deposition vesicle (the within which membrane bound organelle where silica deposition occurs)[48]. Therefore, using fluorescence microscopy to determine the intracellular location of these kinases would be a quick screening method to identify those that reside within the endomembrane system making them interesting candidates for silaffin and/or silacidin phosphorylation. Furthermore, using the biochemical methods (i.e. differential centrifugation, sucrose density gradient centrifugation) established in this study, a proteomic analysis of the ER and Golgi membrane fractions can be used to identify any membrane associated kinases that phosphorylate biomineral proteins. The proteomic study of the ER and Golgi membrane fractions may also lead to the identification of other proteins such as glycosyl and methyl transferases, peptidases, and hydrolases that posttranslationally modify silaffins.

It should be noted that the majority of the silacidin phosphorylation activity was observed predominantly in the Golgi fractions, along with the majority of casein kinase activity. In addition, the sequence analysis program NetPhosK 1.0[131] predicted 54

putative casein kinase phosphorylation sites in recombinant silacidin, with greater than 60 % probability for many of the sites. Though purely speculative, there may exist kinases analogous to the hydroxapatite biomineralization kinases found in mammalian ER and Golgi fractions[74-76] that are involved in phosphorylation of the diatom biomineralization protein, silacidin. A search of the diatoms *T. pseudonana* and *P. tricornutum* genome databases[85-86] for casein kinases revealed 15 putative kinases (tp38011, tp22118, tp21996, tp17468, tp17208, pt11285, pt14627, pt12614, pt7026, pt48012, pt54091, pt42322, pt21821, pt51110, and pt4216) with significant sequence similarity (e values < 2e-10) to more than 40 proven or predicted casein kinases from a broad range of organisms (including other biomineralizing organisms). There were no putative N-terminal signal peptides predicted for any of the putative kinases by the Signal P program[109] nor was the C-terminal ER retention signal DDEL present in any of these sequences. However, ~ 75 % of the casein kinase-like protein sequences appeared to be incomplete.

4.4 Assessing the Importance of Diatom Biomineralization Proteins

While the *in vitro* characterization of ER and Golgi kinases from diatoms, including tpSTK1, may implicate them in diatom biomineralization, the function and importance of these proteins *in vivo* may differ. Unfortunately, methods typically used to determine the role of a particular protein *in vivo* (i.e. gene knockdown or gene knockout) have yet to be established in the diatom *T. pseudonana*. However, there are alternative methods. 1) Biomineralization kinases can be (presumably specifically) inhibited in vivo using small organic molecules. With more than 200 kinase inhibitors available, one of

them may be a great inhibitor of tpSTK1 activity (as judged by low IC_{50} values and kinase specificity). Also, if there are hydroxyapatite biomineralization-like kinases present in diatoms that are indeed isoforms of casein kinases I and II, then these kinases can specifically be inhibited by the compounds D4476 and IQA, respectively[132-133]. In addition to chemical inhibition, antibody based inhibition has been used to suppress the activity of a protein[42,72]. In this study, anti-tpSTK1 IgGs inhibited tpSTK1 kinase activity on recombinant silaffins *in vitro*. 2) Another approach for determining whether tpSTK1 (or other future diatom biomineralization kinases) phosphorylates biomineral forming proteins *in vivo* take advantage of the established genetic manipulation techniques for *T. pseudonana*. TpSTK1 mutants (such those with an altered ATP binding site) could be generated and introduced into *T. pseudonana* to compete with the endogenous protein for substrates, or tpSTK1 can be overexpressed in the cells. Regardless of the approach used, it would be interesting if the any of these experiments result in a change in the degree of silaffin (or silacidin) phosphorylation that would cause a phenotypic change in the structure of diatom biosilica.

APPENDIX A: AMINO ACID SEQUENCES OF RECOMBINANT SILAFFINS

rSilC

```
  1 MSSKKSGSYYSYGTKKSGSYSGYSTKKSASRRILSSKKSGSYSGYSTKK
 50 SGSRRILSSKKSGSYSGSKGSKRRILSSKKSGSYSGSKGSKRRNLSSKKS
100 GSYSGSKGSKRRILSSKKSGSYSGSKGSKRRNLSSKKSGSYSGSKGSKR
149 RILSGGLRGSMHHHHHH
```

rSilN

```
  1 MAAQSIADLAAANLSTEDSKSAQLISADSSDDASDSSVESVDAASSDVS
 50 GSSVESVDVSGSSLESVDVSGSSLESVDDSSEDSEEEELRILHHHHHH
```

rSil3

```
  1 MEGHGGDHSISMSMHSSKAEKQAIEAAVEEDVAGPAKAAKLFKPKA
 47 SKAGSMPDEAGAKSAKMSMDTKSGKSEDAAAVDAKASKESHMSISG
 92 DMSMAKSHKAEAEDVTEMSTAKAGKDEASTEDMCMPFAKSDKEMS
137 VKSKQGKTEMSVADAKASKESSMPSSKAAKIFKGKSGKSGSLSMLK
183 SEKASSAHSLSMPKAEKVHSMSAHHHHHH
```

rSil1L

```
  1 MVPGLTEMPTISPTEDHYFFGKSHKSHKSHKSKATKTLKVSKSGKSAKS
 50 SKSSGRRPLFGVSQLSEGIAVGYAKSSGRSSQQAVGSWMPVAAACILG
 98 ALSFFLNHHHHHH
```

Figure A.1 Amino acid sequences of recombinant silaffins. Hydroxyl amino acids (serine S, threonine T, and tyrosine Y) are in red.

APPENDIX B: MASS SPECTRA OF RECOMBINANT SILAFFINS

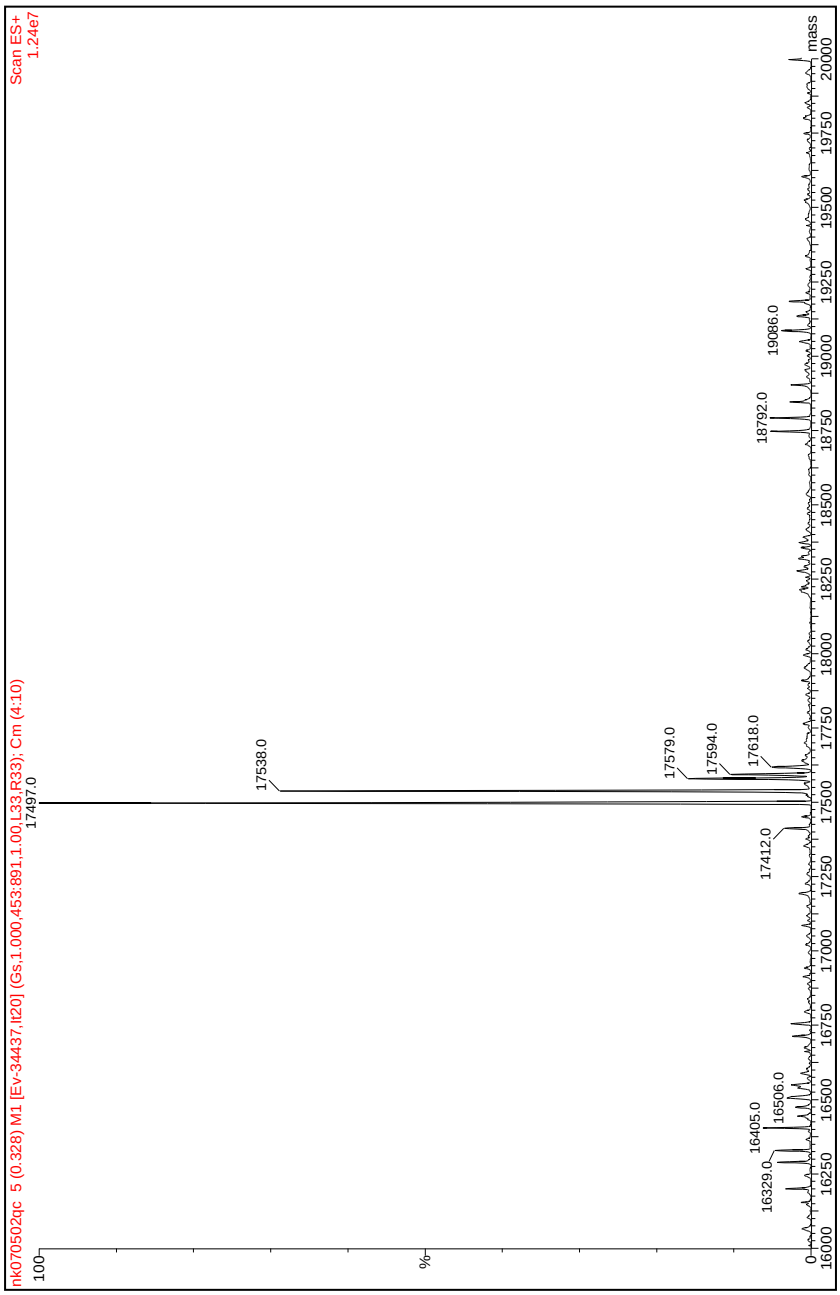


Figure B.1. Mass spectrum of recombinant silaffin rSilC. The expected molecular mass is 17.625 kDa. The observed molecular mass of 17.494 kDa corresponds to the loss of the start methionine, which can be cleaved off during expression of the protein in *E. coli*.

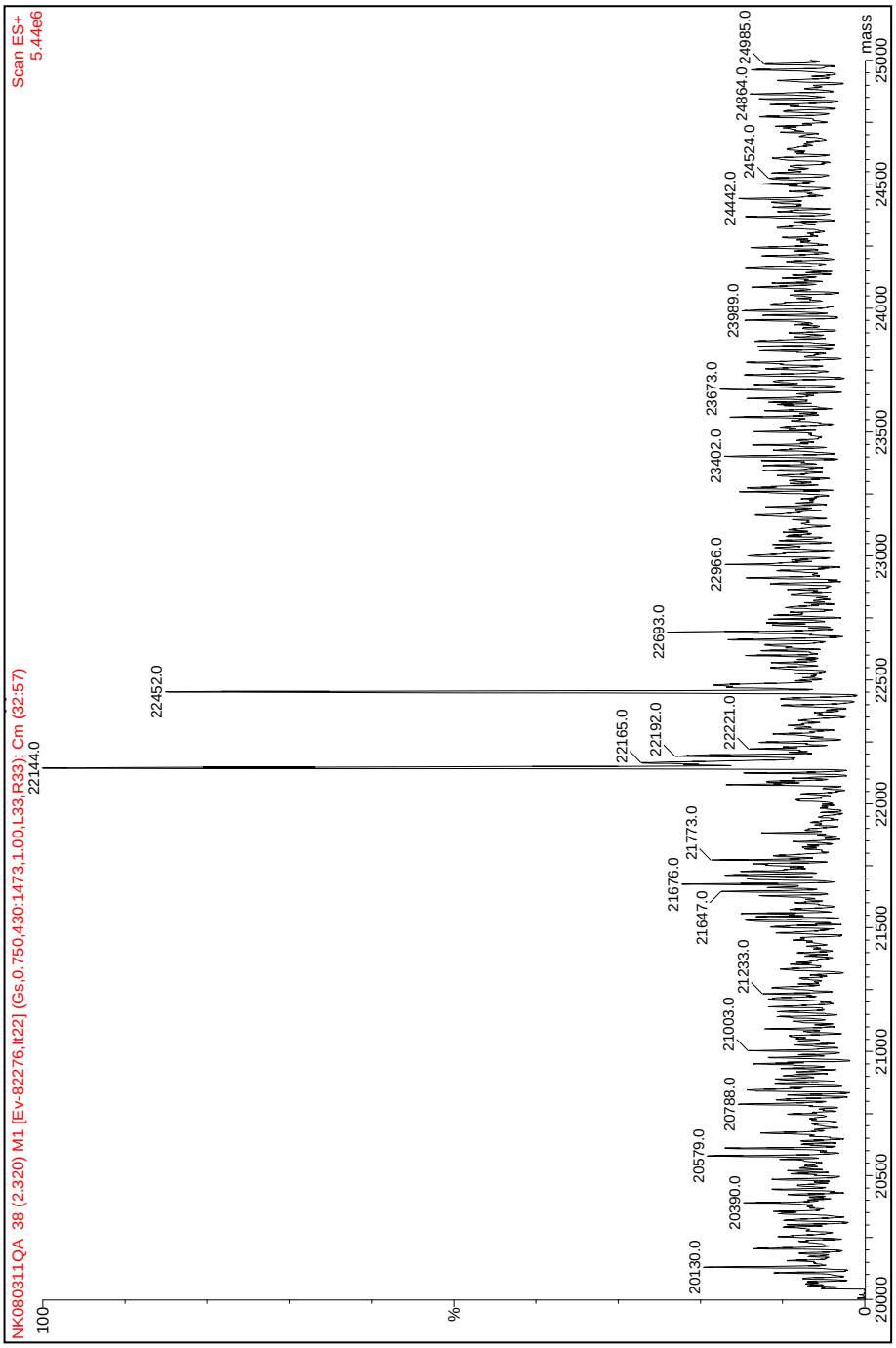


Figure B.2. Mass spectrum of recombinant silaffin rSil3. The molecular mass that is observed in the spectrum corresponds to the expected molecular mass of 22.1 kDa.

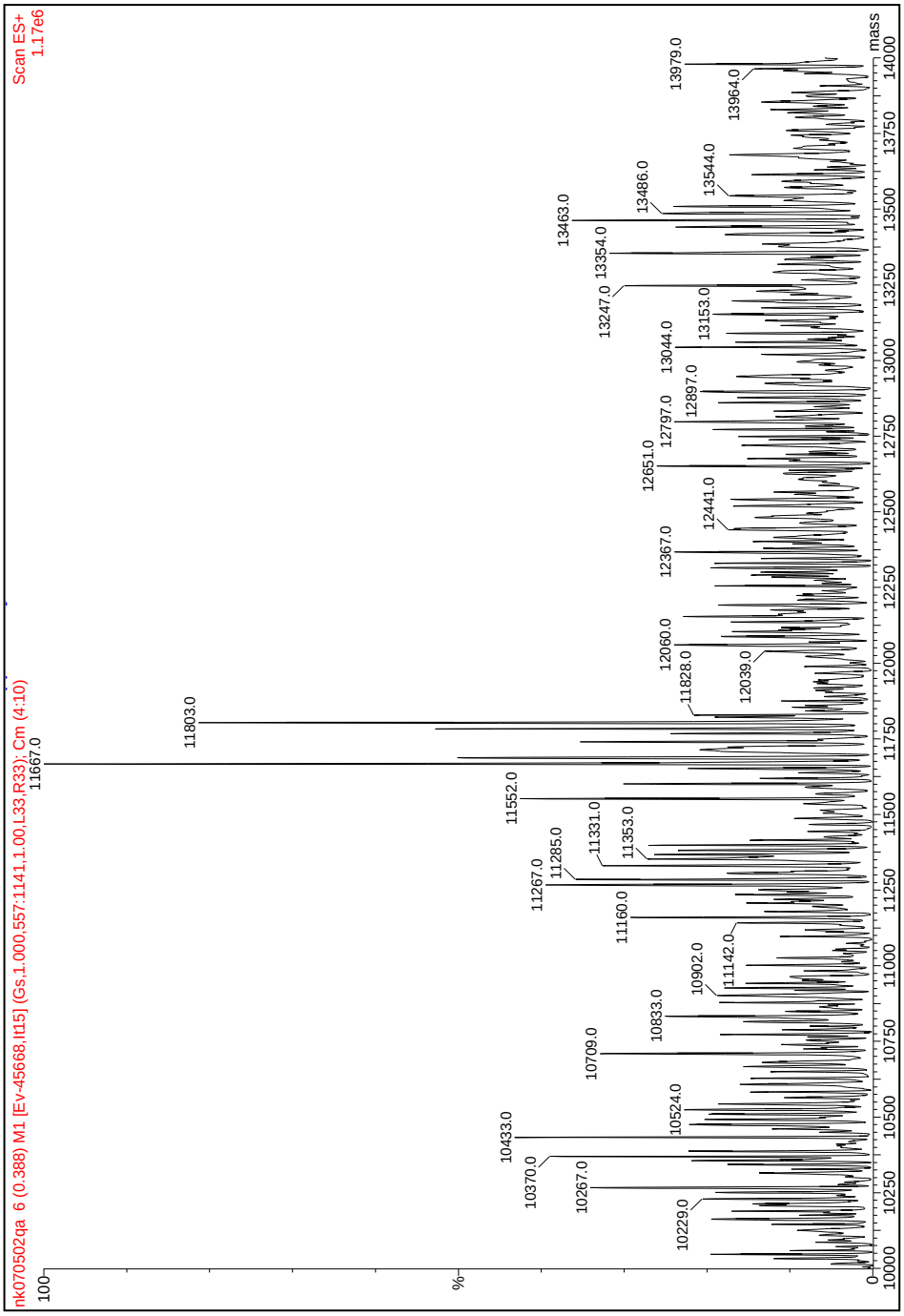


Figure B.3. Mass spectrum of recombinant silaffin rSil1L. The expected molecular mass is 11.8 kDa. The observed molecular mass of 11.667 corresponds to the loss of the start methionine, which can be cleaved off during expression of the protein in *E.coli*.

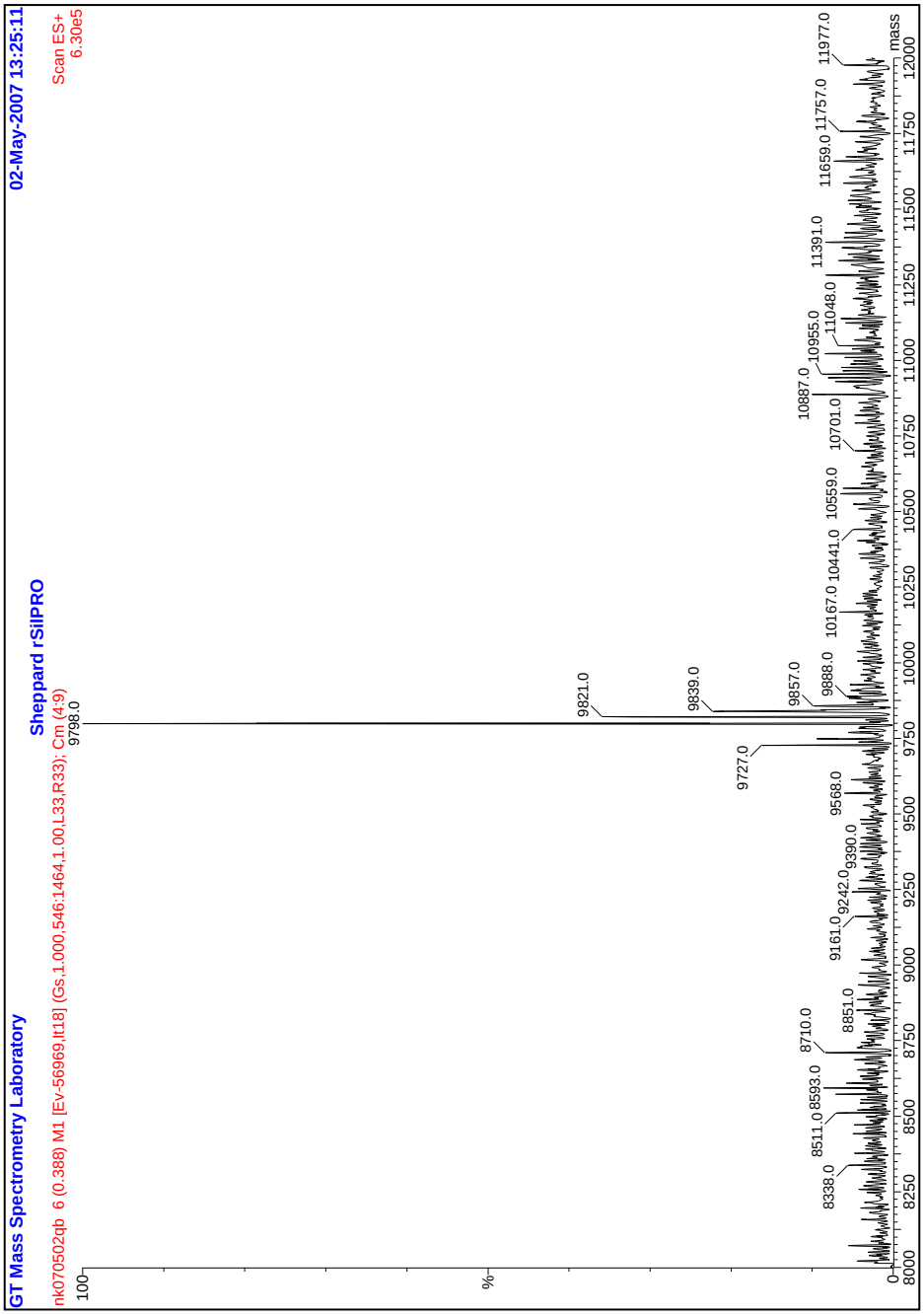


Figure B.4. Mass spectrum of recombinant silaffin rSilN. The expected molecular mass is 9.978 kDa. The observed molecular mass of 9.799 kDa corresponds to the loss of the start methionine, which can be cleaved off during expression of the protein in *E. coli*.

APPENDIX C: *tpak1* GENE SEQUENCE

```
1    ATGACAGCTACACCGTATAACAGGGAGGAAGTGCAGAGGAGCCTTCGTGAAAAGTAAGTGTCGACAAACA
      M  T  A  T  P  Y  N  R  E  E  V  Q  R  S  L  R  E  K

71   GCGAATGAGCAGTATCCTATCAAGGTGTGGTCGCCGGTAGACGAGGACAAGCCTCGCGTGTGAACGCTTT


141  TCGTGGTCTTTATTTCGCCCATTGCTCAAGGTGCCCTGTTGCTACTTCTATGCTTCTGTTCTCAGAATCT
                                                                        I

211  GGAGAGCTGCTGCCACGGCCGTATCAGAGGAAGATCCATGGAAGAAACATGACATAACATCAATCCCTGG
     W  R  A  A  A  T  A  V  S  E  E  D  P  W  K  K  H  D  I  T  S  I  P

281  TGAGTACTGATTGTATTCCACTCTTGCTCTACCTACAACAAACTTTTTCTAACATGGTCTCTTGCGTTGC


351  TCACAGCCGAACGTGTAGTTCGACACTGTTACAACCCCAACACTCGCCAGTTTAAAAAGGATGAGACGAG
          A  E  R  V  V  R  H  C  Y  N  P  N  T  R  Q  F  K  K  D  E  T

421  TGAGTTCATCAGTTTAAGGCAATGTATCAATTGCAATTGTATTATGGCTTTATCACATATTATTCACTCC


491  TTGTTGGACTAACAGTCGTCAAGGTGGAGAAGGAACCTTTCACTCACGGGGCGATGCGACATTGTTTCAG
                    I  V  K  V  E  K  E  P  F  T  H  G  A  M  R  H  C  F  R

561  ATTGAAGAAGCTAGCAACTCCACCACAATCATCATCCAACCATCGATTCCACAGTTATGGATGGAGTAGG
       L  K  K  L  A  T  P  P  Q  S  S  S  N  H  R  F  H  S  Y  G  W  S  R

631  GCGTTGAACTATGTTGCAAAGTGCTATCTAAAAGAAGACGGGACAATCGATACGAGCCGAAAGGGAACAG
      A  L  N  Y  V  A  K  C  Y  L  K  E  D  G  T  I  D  T  S  R  K  G  T

701  ACAACGTACTGACGGATATTATTCTACAATACGAAGCTGCCCATTGGTCGGCCCTCTTTAACGAAGCTAA
     D  N  V  L  T  D  I  I  L  Q  Y  E  A  A  H  W  S  A  L  F  N  E  A  N

771  TCCTCCTAGAAAGATTGATTTCATTAGAGCTTATGCCATGGAGTTTGTGGATCGTCCAGGCAAACCAATG
       P  P  R  K  I  D  F  I  R  A  Y  A  M  E  F  V  D  R  P  G  K  P  M

841  TTTGCAGTTGAGCGCTACATTGCGGGGAACGATAGCTACGGATGTGGTTTCCACAAACACAATACCAACT
      F  A  V  E  R  Y  I  A  G  N  D  S  Y  G  C  G  F  H  K  H  N  T  N

911  CTGGTTTCATTGACTTGGAGATACGCCGGAAGACTCCTCAGGTGTTTTCAGCTCATAGCTTCTACGCTTC
     S  G  F  I  D  L  E  I  R  R  K  T  P  Q  V  F  S  A  H  S  F  Y  A  S

981  GGAAGGGACGAGACTCGTTGCAGATGTTCAGGGCGTGGGAGATCTTTATACTGATCCGCAGGTGCTATCC
       E  G  T  R  L  V  A  D  V  Q  G  V  G  D  L  Y  T  D  P  Q  V  L  S

1051 ATTGACTATCGCTTTGGTGACGGAGATCTCGGCCCTAGAGGAATGGCTTTGTTCTTCAAAACCTTTCGTC
      I  D  Y  R  F  G  D  G  D  L  G  P  R  G  M  A  L  F  F  K  T  F  R

1121 ATTGTGATATGAGCGACCTTCTGGGGATTCCAATCTTTCCACTATCAAGAAATGAACGTCGCCACCAGGC
     H  C  D  M  S  D  L  L  G  I  P  I  F  P  L  S  R  N  E  R  R  H  Q  A

1191 TAAATACGATGACGAAGAAAGCACTTTAAGTGAAGCGTCATCTGCTGTAGAGGAAGATCTAAGGTGCAGA
       K  Y  D  D  E  E  S  T  L  S  E  A  S  S  A  V  E  E  D  L  R  C  R

1261 TTCAAGATGCTAGATGCCAATAGGCAACGGAGGAAAACAGTCTTAATGAGACCGATTGACATCATGCAAT
      F  K  M  L  D  A  N  R  Q  R  R  K  T  V  L  M  R  P  I  D  I  M  Q
```

```
1331 CAGAGGACAAGGCCGACACCGCGAAAAGGTCTAATATTTCCAACGTCTCGTTGACGATTCGTCAATCCAT
     S  E  D  K  A  D  T  A  K  R  S  N  I  S  N  V  S  L  T  I  R  Q  S  M

1401 GAGACAACTGAACATACAAAAGTCAGTGATTCATCGTTCAAAATCGGATGTCGATGAGATCACATCCTGC
       R  Q  L  N  I  Q  K  S  V  I  H  R  S  K  S  D  V  D  E  I  T  S  C

1471 TTGCAAATGGCAGTGACTGATGCGGTATTTGATCACACAGCGTTTCATAGATACGAATCTGGAGAACTGA
      L  Q  M  A  V  T  D  A  V  F  D  H  T  A  F  H  R  Y  E  S  G  E  L

1541 AACCTCGTCACTTCCACGGGACACCGAAGCATCAAACTCAGTTAACCGTGGTACCAGTGAAAGCATCGCC
     K  P  R  H  F  H  G  T  P  K  H  Q  T  Q  L  T  V  V  P  V  K  A  S  P

1611 GCCAATGAAGATAACTGACGAGACAAAGGCAAATCTCGGCAAAGTAAGTATTGTATCGCGTTATTGGCAT
       P  M  K  I  T  D  E  T  K  A  N  L  G  K

1681 TTACTGTGTGTGTTATTTAGTACAGTGATTTGTCGCCGAACTAGCCATCTCACTCTTATCTCGTGTCTAC


1751 GCAGGTCCACTACCACTTAGCCTGTCTTCACGGGCTAGATCGCTTCCCAGAGATTGTTCCTACCAATGCG
          V  H  Y  H  L  A  C  L  H  G  L  D  R  F  P  E  I  V  P  T  N  A

1821 TCCTCTGGCGTGGAGGATCTTCCCAGCCATGATGTCTTCTCTGTGGTGTTTCATCTCTCTCATGCAGCGT
      S  S  G  V  E  D  L  P  S  H  D  V  F  S  V  V  F  H  L  S  H  A  A

1891 CACTCTACAACGTACCAGCTTGTCTCTCATTGGCTCGTGCTCTAATTGGCCTTGATAGCTCGGTATCTCC
     S  L  Y  N  V  P  A  C  L  S  L  A  R  A  L  I  G  L  D  S  S  V  S  P

1961 TTTGCTCAAAGCGAACGTCCGTATTGATTTTGAGCTATCAAAGGATTTGTGCTGGCGAGGGATGACTGCT
       L  L  K  A  N  V  R  I  D  F  E  L  S  K  D  L  C  W  R  G  M  T  A

2031 CAAAAGGCCCGTGCTGCTCCGAAAGTTGCGGCTGGTTGTTTGTTGTATCAAATCCTTGAAGACGAAGGTA
     Q  K  A  R  A  A  P  K  V  A  A  G  C  L  L  Y  Q  I  L  E  D  E  G

2101 CAACAGGCGATGTAGAAAAGATGAACATTCTAGAGGAGACGCTGAACCAGATGAAAGTTGCTACCGAAGA
     T  T  G  D  V  E  K  M  N  I  L  E  E  T  L  N  Q  M  K  V  A  T  E  E

2171 GGAAGCCATTTTGAAGGAGCATGCAAACAAATTAACCCGAGGTAAGGCCGCAGGCTTCCACGTGGGCGAC
       E  A  I  L  K  E  H  A  N  K  L  T  R  G  K  A  A  G  F  H  V  G  D

2241 AAAGTGGAAGGAAACTACTTCATGGAAGGAACATTCTACAGTGGAGTTGTGGTTGAAGTTCTCGATGGTG
     K  V  E  G  N  Y  F  M  E  G  T  F  Y  S  G  V  V  V  E  V  L  D  G

2311 GCAACTCTGTTGTGGTTCAATACGACGACGATGGATCGGCTGAACAATTGACGATTGACAATGTCCGCTC
     G  N  S  V  V  V  Q  Y  D  D  D  G  S  A  E  Q  L  T  I  D  N  V  R  S

2381 TATTGAACCTGCAACGGCCATTACTGATGATCAGAACTCTCCCTTGTCAGATGAAGAAGCATTGGGCAGC
       I  E  P  A  T  A  I  T  D  D  Q  N  S  P  L  S  D  E  E  A  L  G  S

2451 GTTAACACTGATGAAGAGTGCTTGTTTGAAGACTATGAGTTGATGGCTAAACTCGCGGACTTGAAAGTGA
      V  N  T  D  E  E  C  L  F  E  D  Y  E  L  M  A  K  L  A  D  L  K  V

2521 AGGCTGGAAAGCGAAACGATGCGCGGGTGCTATATCAGGATGCAGCAGAGCTTGCCATGTCTGCGGGTAA
     K  A  G  K  R  N  D  A  R  V  L  Y  Q  D  A  A  E  L  A  M  S  A  G  K

2591 GATGAAGACGGCGAATTGCTTGAGTATGAAGGCCGCCCAACTTGAAGGTTAG
       M  K  T  A  N  C  L  S  M  K  A  A  Q  L  E  G  *
```

Figure C.1 DNA sequence of the *tpak1* gene. Introns are indicated in grey italics. The deduced amino acid sequence appears below the DNA sequence.

APPENDIX D: AMINO ACID SEQUENCE OF RECOMBINANT TpAK1

1 MTATPYNREEVQRSLREKIWRAAATAVSEEDPWKKHDITSIPAERVVRHC

51 YNPNTRQFKKDETIVKVEKEPFTHGAMRHCFRLKKLATPPQSSSNHRFHS

101 YGWSRALNYVAKCYLKEDGTIDTSRKGTDNVLTDIILQYEAAHWSALFN

150 EANPPRKIDFIRAYAMEFVDRPGKPMFAVERYIAGNDSYGCGFHKHNTNS

200 GFIDLEIRRKTPQVFSAHSFYASEGTRLVADVQGVGDLYTDPQVLSIDYRF

251 GDGDLGPRGMALFFKTFRHCDMSDLLGIPIFPLSRNERRHQAKYDDEEST

301 LSEASSAVEEDLRCRFKMLDANRQRRKTVLMRPIDIMQSEDKADTAKRS

350 NISNVSLTIRQSMRQLNIQKSVIHRSKSDVDEITSCLQMAVTDAVFDHTAF

401 HRYESGELKPRHFHGTPKHQTQLTVVPVKASPPMKITDETKANLGKVHY

450 HLACLHGLDRFPEIVPTNASSGVEDLPSHDVFSVVFHLSHAASLYNVPAC

500 LSLARALIGLDSSVSPLLKANVRIDFELSKDLCWRGMTAQKARAAPKVA

549 AGCLLYQILEDEGTTGDVEKMNILEETLNQMKVATEEEAILKEHANKLTR

599 GKAAGFHVGDKVEGNYFMEGTFYSGVVVEVLDGGNSVVVQYDDDGSA

646 EQLTIDNVRSIEPATAITDDQNSPLSDEEALGSVNTDEECLFEDYELMAKL

697 ADLKVKAGKRNDARVLYQDAAELAMSAGKMKTANCLSMKAAQLEGH

743 HHHHH

Figure D.1 Amino acid sequence of recombinant tpAK1. The C-terminal hexahistidine tag is indicated in blue.

APPENDIX E: TpAK1 PHOSPHORYLATION ACTIVITY

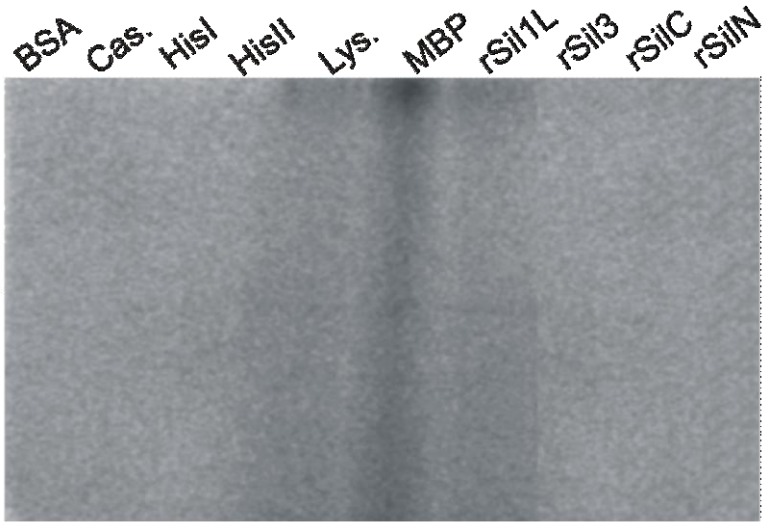


Figure E.2. Recombinant tpAK1 phosphorlation activity. SDS-PAGE analysis of recombinant tpAK1 phosphorylation activity in the presence of [γ-^{32}P]ATP using recombinant silaffins (rSilN, rSilC, rSil3, and rSil1L) and commercial proteins as substrates. Lys.- Lysozyme, His-Histone, MBP- Myelin Basic Protein, Cas.- dephosphorylated casein, BSA- bovine serum albumin.

APPENDIX F: *tpstk1* GENE SEQUENCE

```
   1 ATGAGAGTCATACGCAATGTTTCTCTCGTAGTCATCGTCACGTCATGCACCTTTCTTGTGTCCGTCACCT
     M  R  V  I  R  N  V  S  L  V  V  I  V  T  S  C  T  F  L  V  S  V  T

  71 CCTTCACATCACAACGCACCGTCACCTCGGCATCTCCATCCTACACCAACTCAAACAACAATCTCTACTC
     S  F  T  S  Q  R  T  V  T  S  A  S  P  S  Y  T  N  S  N  N  N  L  Y  S

 141 GCCACCGCTCTTTGCATCAGGTCCACCCGTTATCAGCAACTGGCGTTACCAATCCTCTTCTGGTACCATT
       P  P  L  F  A  S  G  P  P  V  I  S  N  W  R  Y  Q  S  S  S  G  T  I

 211 TCAGGAACAGTCACAAACCATCCCACCATTCCTGATGGAGACAGTATCACTACGTCACCAGTGAGGGATG
     S  G  T  V  T  N  H  P  T  I  P  D  G  D  S  I  T  T  S  P  V  R  D

 281 GAAATGGCATCAACGAGAATACAATAGTGCAGACGAAGAGTGGGAGCAAGTATAAGTTGAGCAATGTTGC
     G  N  G  I  N  E  N  T  I  V  Q  T  K  S  G  S  K  Y  K  L  S  N  V  A

 351 TTGGGGAGCGACGATTCCTGCAGTGGAGACGAAGAAGGCGATGCAAAAGGAAGCTGAGGCTGCCGAGAAG
       W  G  A  T  I  P  A  V  E  T  K  K  A  M  Q  K  E  A  E  A  A  E  K

 421 GCGAGGAAGGTCGAGGCGGAGAAAGCAAAAACTGCAGCTGCCAAGGCGAAAACAGCATCCAAAGCAACAC
     A  R  K  V  E  A  E  K  A  K  T  A  A  A  K  A  K  T  A  S  K  A  T

 491 CAAAAGCAACATCACCGGCTGCAGCTACTACCAAACAACCATCCTCTGCAGAACTGCGTCGACTTGCCAA
     P  K  A  T  S  P  A  A  A  T  T  K  Q  P  S  S  A  E  L  R  R  L  A  K

 561 ATCAACATTCTCTCTCACCGGTAACACAGTCGGCTCCAACGGGAAATACCTCCTCGCAGGCAAACCCCGT
       S  T  F  S  L  T  G  N  T  V  G  S  N  G  K  Y  L  L  A  G  K  P  R

 631 CAATCTTCCGGCAGAGCTGCCAAAATCTGGACTGCGTACATTGCCGACCCCATCGAGCCTGGAGTACCAC
     Q  S  S  G  R  A  A  K  I  W  T  A  Y  I  A  D  P  I  E  P  G  V  P

 701 TTGGTTTTAGCGAGGGAGACGAGTCAAAAGTACAACCACTCACAATCAAATTAAGTCCAGATGTCGATAG
     L  G  F  S  E  G  D  E  S  K  V  Q  P  L  T  I  K  L  S  P  D  V  D  R

 771 ATTGTACTTGGAGAATGCAAACTACAACAAAGTTCAATCGGGATTGTTCATGGGGAGATTTGTAAAGAAG
       L  Y  L  E  N  A  N  Y  N  K  V  Q  S  G  L  F  M  G  R  F  V  K  K

 841 ATGGACTACATTGAGAATGTACCATCGAGCGATCGTTCTTTGGATGGGAAAGTGTCAGCTTTAATTATTG
     M  D  Y  I  E  N  V  P  S  S  D  R  S  L  D  G  K  V  S  A  L  I  I

 911 AGTCGGGGCAATATGATCTCAAGGCGTTGTTGTCGGCGAGGAAGGGACTTCCCTTGGCAGGGAAAGCATT
     E  S  G  Q  Y  D  L  K  A  L  L  S  A  R  K  G  L  P  L  A  G  K  A  L

 981 GAGGGATGCCGCAGCTGCAGCAGGGCAGTGCATCCAAGCAGTACATTCCAGCAACCTCGTTTGGACTGAT
       R  D  A  A  A  A  A  G  Q  C  I  Q  A  V  H  S  S  N  L  V  W  T  D

1051 TTGAAGACGGAGAATTTCATTGTCGTGCGGAACGGTAACTTGTATCAAGAAGTGGAGGATGTCAAAGGTT
     L  K  T  E  N  F  I  V  V  R  N  G  N  L  Y  Q  E  V  E  D  V  K  G

1121 CAATGGGGCTGCCTGGAGTGAAGGGCATTGATTTGGAGTCAGTGATTGCCAAGGGGGGCAATCCAATCGA
     S  M  G  L  P  G  V  K  G  I  D  L  E  S  V  I  A  K  G  G  N  P  I  D

1191 CTACTCGCCCGAGGCATGTCCTCCAGAGTTTGCCAAAGCGTACATGGAAGGAATCGGGGGTGATTTCGTC
       Y  S  P  E  A  C  P  P  E  F  A  K  A  Y  M  E  G  I  G  G  D  F  V

1261 TTGGACTATTCATACGATATGTGGAGTCTGGGAATGATGCTGTACGAACTTTCCACTGGCAAAGCATACT
     L  D  Y  S  Y  D  M  W  S  L  G  M  M  L  Y  E  L  S  T  G  K  A  Y

1331 TTGACAAAAAGGGTCCAAGTCAGGTGACGAAGTTATTGTGTACTGATACATTTGAGGCAGATGTGAGTGC
```

```
     F  D  K  K  G  P  S  Q  V  T  K  L  L  C  T  D  T  F  E  A  D  V  S  A

1401 TGTGCCTGATGAAAAGTTACGAGATTTGATTGGGAACTGCTTGAACTTGGATCCTCGAAAGCGACCGGAC
       V  P  D  E  K  L  R  D  L  I  G  N  C  L  N  L  D  P  R  K  R  P  D

1471 ATTACTGGATTCTTGTTGCATCCGTACTTTCTTCTCAGCGGCATCGGTCCTATCAGCTTTTAG
       I  T  G  F  L  L  H  P  Y  F  L  L  S  G  I  G  P  I  S  F  *
```

Figure F.1 DNA sequence of the *tpstk1* gene. The deduced amino acid sequence appears below the DNA sequence.

APPENDIX G: AMINO ACID SEQUENCE OF RECOMBINANT TpSTK1

1 HHHHHH**DYDIPTT**ENLYFQGMRVIRNVSLVVIVTSCTFLVSVTSFTSQRT

51 VTSASPSYTNSNNNLYSPPLFASGPPVISNWRYQSSSGTISGTVTNHPTIPD

103 GDSITTSPVRDGNGINENTIVQTKSGSKYKLSNVAWGATIPAVETKKAMQ

153 KEAEAAEKARKVEAEKAKTAAAKAKTASKATPKATSPAAATTKQPSSAE

202 LRRLAKSTFSLTGNTVGSNGKYLLAGKPRQSSGRAAKIWTAYIADPIEPG

252 VPLGFSEGDESKVQPLTIKLSPDVDRLYLENANYNKVQSGLFMGRFVKK

301 MDYIENVPSSDRSLDGKVSALIIESGQYDLKALLSARKGLPLAGKALRDA

351 AAAAGQCIQAVHSSNLVWTDLKTENFIVVRNGNLYQEVEDVKGSMGLP

399 GVKGIDLESVIAKGGNPIDYSPEACPPEFAKAYMEGIGGDFVLDYSYDMW

449 SLGMMLYELSTGKAYFDKKGPSQVTKLLCTDTFEADVSAVPDEKLRDLI

498 GNCLNLDPRKRPDITGFLLHPYFLLSGIGPISFHHHHHH

Figure G.1 Amino acid sequence of recombinant tpSTK1. The N- and C- terminal hexahistidine tags are indicated in blue. The spacer region is in bold and the rTEV protease cleavage site is in red.

APPENDIX H: SILACIDIN AMINO ACID SEQUENCE

1	MVKYNVLAFIAVLGVSLINTSSA	
24	KTSLRGHRQLAKPEKLGNANTSYALGSSINKVRRL	
58	SSSEDSGDSPPSDESEESEDSVSSEDEDRRL	Silacidin A
89	SSSEDSGDSPPSDESEESEDSVSSEDEDRRL	Silacidin A
120	SEDSVDSLPSDESEESEDSVSSEDEDRRL	Silacidin C
149	SEDSGDSLPSDESEESEDSVSSEDEDRRL	Silacidin B
178	SSSEDSGDSPPSDESEESEDSVSSEDEDRRL	Silacidin A
209	SSSEDSGDSPPSDESKESGDSVSSED	Silacidin A

Figure H.1. Amino acid sequence of silacidin. The spacer sequence RRL between individual silacidin repeats (Silacidin A, B, or C) is shown in green. Hydroxyl amino acids are in red. The N-terminal signal sequence is boxed.

APPENDIX I: AMINO ACID SEQUENCE OF RECOMBINANT rSILN$_{H10}$rSIC

1 MAAQSIADLAAANLSTEDSKSAQLISADSSDDASDSSVESVDAASSDVSG

51 SSVESVDVSGSSLESVDVSGSSLESVDDSSEDSEEEELRILHHHHHHHHHH

102 **Met**INKVRRLSSSEDSGDSPPSDESEESEDSVSSEDEDRRLSSSEDSGDSPPS

153 DESEESEDSVSSEDEDRRLSEDSVDSLPSDESEESEDSVSSEDEDRRLSEDS

205 GDSLPSDESEESEDSVSSEDEDRRLSSSEDSGDSPPSDESEESEDSVSSEDE

257 DRRLSSSEDSGDSPPSDESKESGDSVSSED

Figure I.1. Amino acid sequence of recombinant rSilN$_{H10}$rSic fusion protein. The methionine between rSilN$_{H10}$ and rSic is shown in green bold. The decahistidine tag is indicated in blue. Hydroxyl amino acids are in red.

APPENDIX J: *tpstk2* GENE SEQUENCE

```
1    ATGCCCTTCTGTGGCAGGTGGATTGCAAGCTACTACTCACCTTGTGTTCCATCTCGCCCTACGCATGAGT
      M  P  F  C  G  R  W  I  A  S  Y  Y  S  P  C  V  P  S  R  P  T  H  E

71   GGGCTGCAGCCGACAAGAACTTTCAACTGGGCCGGTTAACTACGGACGAAAAGGCAGGCATTACATCAGT
     W  A  A  A  D  K  N  F  Q  L  G  R  L  T  T  D  E  K  A  G  I  T  S  V

141  TCGCACGAAAGATGATTGGGTAGAGAATTACGTCTCTCAAACGGTTGCGACTCGTATCGAGAGGGAAAGA
       R  T  K  D  D  W  V  E  N  Y  V  S  Q  T  V  A  T  R  I  E  R  E  R

211  GCTGATGGCAATCGTCACTTCTACAAAAACAAAGACTGCCAAGAAGCCTTTGCTCGCTTCACGTGTTGGT
      A  D  G  N  R  H  F  Y  K  N  K  D  C  Q  E  A  F  A  R  F  T  C  W

281  AAGTCAGCTACAGTGTGCATTAAGTGAAGAAGTTGCTGTTGTTGTTGTTGTTGTTGTTGTTGTTTCCTTG


351  TGTTGCAACCTGATCTTTCTTTCGTTGTTCTTCCAAAGGCTCAACTTCCCTCGTTGTGACGAACGATTTG
                                          R  L  N  F  P  R  C  D  E  R  F

421  AAGAGTCGCTACCTCTTTGCCGTTCCGCTTGTGAAAATATATTTCGAGTATGTGGCTTTGAGAATGATCT
     E  E  S  L  P  L  C  R  S  A  C  E  N  I  F  R  V  C  G  F  E  N  D  L

491  TTGGAGATGTGAAGAAGATGTGATTGATGGCAATGACGAGTACGATCTGAGAGGATTCTTTCCAGGACAG
       W  R  C  E  E  D  V  I  D  G  N  D  E  Y  D  L  R  G  F  F  P  G  Q

561  CCGTTCAAGAGGAATGAGTTCTTTCCAAAGAGCAACGGAGAGCCAAAGGCAGTTTGTACTCCTAGTATCA
      P  F  K  R  N  E  F  F  P  K  S  N  G  E  P  K  A  V  C  T  P  S  I

631  AAGGCTCGGCACCTTCACGATACGGTGGTGTTGCGTTTGTGGCCGTGATTGGTATCAATGCGATGCTAGT
     K  G  S  A  P  S  R  Y  G  G  V  A  F  V  A  V  I  G  I  N  A  M  L

701  ACTTAGTCTTTGCTCTACCGTTTAGACTCAACATGTATTAGCTACTAGCACGTAATGATCAATCTACTAT

771  TGAATATGAGATTTATCTCGTTTCGTAACTCAAGAAATGAGTTCGAGATTACTGTGCACTTCACCTCCAA

841  ACAATGAGTGTGACGAGCGACGTTTGAACTAGAAGTCAAAGTGATTCCAATAAGTACCAGTATCTGTGAC

911  CTACCATAAAGGATTCTTGATCTTTATGGTATCCGTATATTCATGATCGAAGGCAAGCTAGGTAAATCCA

981  ACAGTATTAACTCATCGTAGTGTGAATTGAATATGGTGCAATGACATCGACGTAGCTAGTGGTGCAACGT


1051 CGGTGAGAAGAGAGAGGGTGCGTATCCGTTTGATAGTTTGTTAGTAATCTCCTTCAATCACAATCCTCAA

1121 CCAACTCGAAGCCGAATCGAAGCGAAGCGCTTGGAACCATGGCAAGGCGTGGCCTTGCAATAAGCTTGAT

1191 ATACATCTTTCATACAGTCTCATTCACACAGTCATTGATACCATCCTCGCGCTATTGGGTAGGTGACATA
                                        S  L  I  P  S  S  R  Y  W  V  G  D  I

1261 GGCAAAAGTAGAGACACAACTACCCCAACCACAACATCATCAAACCATCACCTGTACCAACGACGACACC
      G  K  S  R  D  T  T  T  P  T  T  T  S  S  N  H  H  L  Y  Q  R  R  H

1331 AATCATCAAGAACAAAACTCTCTGCCTTCTTCCTCAACGACAAAGACGACAATAACAAGACGCCTTCCAA
     Q  S  S  R  T  K  L  S  A  F  F  L  N  D  K  D  D  N  N  K  T  P  S  K

1401 GAATAATGCTACGCCATTGGGGAAGGACGGCTCCTCTGCCAACGCAGACCTCCTCCAAACACTAGAGCTA
       N  N  A  T  P  L  G  K  D  G  S  S  A  N  A  D  L  L  Q  T  L  E  L
```

```
1471 AAAGATGAAGCCCTCTACCAAGCACAAACGGCGGTGTCAAGTCTCGAGAAGGCACTGGAATCTGCTGTTA
      K  D  E  A  L  Y  Q  A  Q  T  A  V  S  S  L  E  K  A  L  E  S  A  V

1541 CGAATTTAGAGAATATGCAACAGCAGTTGCAATTACGGGTGCTTCGTTTGGAGAAGGAATTACGTACTAC
     T  N  L  E  N  M  Q  Q  Q  L  Q  L  R  V  L  R  L  E  K  E  L  R  T  T

1611 CAAGGGTGAATTGACTGATACGATGGGAGAATTACAAAAGACACGGACAGAGTTGCAAAGTACGAGAGAG
       K  G  E  L  T  D  T  M  G  E  L  Q  K  T  R  T  E  L  Q  S  T  R  E

1751 GAGTTGACACAATCGAGAGAGGAACGTGATGGGTTGGACTGGGCATTGGGACAGAGTCAAGAGGGGCAAC
      E  L  T  Q  S  R  E  E  R  D  G  L  D  W  A  L  G  Q  S  Q  E  G  Q

1821 AGAAGGCGGAGACAAGGGTGGAGGAGCTGGAGACGTATTTGGCTACGTTGGGGGTGGACGCTGAGACTAT
     Q  K  A  E  T  R  V  E  E  L  E  T  Y  L  A  T  L  G  V  D  A  E  T  I

1891 TACTGCGAAGAAGAAGGTCGAGGTTAGTGTGATTACTCATACAAGTGTGTTCAGATTGCGATTGACTTGT
       T  A  K  K  K  V  E

1961 CTTACCACTTGACTTCCTGTTTCATTTAGTCAAATCCATGGCAACTGTGGCCGGGGAACTCAAAAACGTC
                                  S  N  P  W  Q  L  W  P  G  N  S  K  T  S

2031 TGTACCAGTACTGAACGATTGGGTAGTAATCAACGGAAGTATCGAGGGAGAGGTTCAAATCTCTGGAAAA
       V  P  V  L  N  D  W  V  V  I  N  G  S  I  E  G  E  V  Q  I  S  G  K

2101 GTCACAAACCACCCTTCCATTCCCGACGGAGATGCCATTGTCACTAGTCCTTTATCGGATGCAACCCAAG
      V  T  N  H  P  S  I  P  D  G  D  A  I  V  T  S  P  L  S  D  A  T  Q

2171 TAGCAGAAAAGAAGATTGTTTCAACCTCATCTGGATCAAAATACAAGCTTGGAAAGCCAATGGATATGCC
     V  A  E  K  K  I  V  S  T  S  S  G  S  K  Y  K  L  G  K  P  M  D  M  P

2241 GGCAAATCAATCTCCATCCAAGTACGTAGGCGGGGGGAAAGGTAGCAATTCACAGCAATACTCAGGATCA
       A  N  Q  S  P  S  K  Y  V  G  G  G  K  G  S  N  S  Q  Q  Y  S  G  S

2311 CGAGCCAGTATTGCACTTCCAGACCTGACTGGTAAAACTATCGGCAATGGACGCTATCTCCTAGCAGGTC
      R  A  S  I  A  L  P  D  L  T  G  K  T  I  G  N  G  R  Y  L  L  A  G

2381 CTGCAACACCTAGTGTGAATGGAAGGAGTTTCATACAAACTGCATACCGATCCACTCCTGTTGGCAAACC
     P  A  T  P  S  V  N  G  R  S  F  I  Q  T  A  Y  R  S  T  P  V  G  K  P

2451 GATTGGCGAGCCACTGGCTATCAAAGTATCGCAAAACAAGGAAGCAATGAAACGTGAGTTTGCCAATTAT
       I  G  E  P  L  A  I  K  V  S  Q  N  K  E  A  M  K  R  E  F  A  N  Y

2521 CAAAAGGTATCAGCAGGCTTGAAGAAGGGGCACTTTATTCGAAGAAATGAGTTTCTACCTGTGGCTGGGA
      Q  K  V  S  A  G  L  K  K  G  H  F  I  R  R  N  E  F  L  P  V  A  G

2591 ATGAAATGCCTGACAAGAGTGCGTTGGTGATGCAACGGGGTGTGGCGGATGTCAAGGCGTTCATGCCAAA
     N  E  M  P  D  K  S  A  L  V  M  Q  R  G  V  A  D  V  K  A  F  M  P  K

2661 AGTAGGGGGAAGACTTGAGGGGGAAATGTTGATGGATTGTGCAGTGACCGCCCTCCGTTGTGTAGAAGCA
       V  G  G  R  L  E  G  E  M  L  M  D  C  A  V  T  A  L  R  C  V  E  A

2731 TTGCATGCTGTGAAGTTGGTGTGGAACGATCTCAAGACGGAGAATTTTGTTGTGATCGAGGACGGAGGTG
      L  H  A  V  K  L  V  W  N  D  L  K  T  E  N  F  V  V  I  E  D  G  G

2801 GTGTGTCGTTCCGAGGAATCGATCTGGAAAGTTGCATGACAGTCAGAACAAACCCAGTGGATTACACTCC
     G  V  S  F  R  G  I  D  L  E  S  C  M  T  V  R  T  N  P  V  D  Y  T  P

2871 AGAAGCATGTCCACCAGAGTTTGCTCAATCATTCCTGGACGGAGATGCAGAGTCGTTCCTTTTAGAGTAC
       E  A  C  P  P  E  F  A  Q  S  F  L  D  G  D  A  E  S  F  L  L  E  Y

2941 TCGTATGACGTATGGAGTTATGGAATGTTCCTCTATGAGATTAGCACCGGTCGTGGCTTCTTTGATGGTT
      S  Y  D  V  W  S  Y  G  M  F  L  Y  E  I  S  T  G  R  G  F  F  D  G
```

```
3011 ACTCAGCGGAGAAAATCACAAAACTGCTTCCATCTTTTGAACCGGACGTGAGTCAAGTACCCGATGCACA
     Y  S  A  E  K  I  T  K  L  L  P  S  F  E  P  D  V  S  Q  V  P  D  A  Q

3081 ACTAGCTGATTTGATTTTGCAGTGCTTGTCAAAGAACCCGAAAGATCGACCGTCGCTTGTCAGAATCGCA
       L  A  D  L  I  L  Q  C  L  S  K  N  P  K  D  R  P  S  L  V  R  I  A

3151 AAACACCCGTACCTCGCAAGTGCGACAGCGAGCAAGACTCCATTTGACTTCTTATTTGGATCATCCATCT
      K  H  P  Y  L  A  S  A  T  A  S  K  T  P  F  D  F  L  F  G  S  S  I

3221 AA
     *
```

Figure J.1 Predicted DNA sequence of the *tpstk2* gene. Introns are indicated in italics. The deduced amino acid sequence appears below the DNA sequence. Sequences that have been confirmed by RACE PCR and RT PCR are red, which is also the putative kinase domain.

APPENDIX K: AMINO ACID SEQUENCE OF RECOMBINANT TpSTK2-KD

1 MDLTGKTIGNGRYLLAGPATPSVNGRSFIQTAYRSTPVGKPIGE

45 PLAIKVSQNKEAMKREFANYQKVSAGLKKGHFIRRNEFLPVA

87 GNEMPDKSALVMQRGVADVKAFMPKVGGRLEGEMLMDCAV

127 TALRCVEALHAVKLVWNDLKTENFVVIEDGGGVSFRGIDLES

169 CMTVRTNPVDYTPEACPPEFAQSFLDGDAESFLLEYSYDVWS

211 YGMFLYEISTGRGFFDGYSAEKITKLLPSFEPDVSQVPDAQLAD

255 LILQCLSKNPKDRPSLVRIAKHPYLASATASKTPFDFLFGSSIHH

300 HHHH

Figure K.1 Amino acid sequence of recombinant tpSTK2 kinase domain. The C-terminal hexahistidine tag is indicated in blue.

REFERENCES

1. Baeuerlein E., ed. 2007. *Handbook of Biomineralization. Biological Aspects and Structure Formation*. Weinheim:Wiley-VCH.

2. Chien, Y. C., Mucci, A., Paquette, J., Sears, S. K., and Vali, H. 2006. "Comparative study of nanoscale surface structures of calcite microcrystals using FE-SEM, AFM, and TEM." *Microsc. Microanal*. **12** (4), 302-310.

3. Zhu, R., Yu, R., Yao, J., Wang, D., and Ke, J. 2008. "Morphology control of hydroxyapatite through hydrothermal process." *J. Alloy Compd.* **457** (1-2), 555-559.

4. Zhou, Z. and Zhao, X. S. 2005. "Opal and inverse opal fabricated with a flow-controlled vertical deposition method." *Langmuir* **21** (10), 4717-4723.

5. Langer, G., Gussone, N., Nehrke, G., Riebesell, U., Eisenhauer, A., Kuhnert, H., Rost, B., Trimborn, S., and Thoms, S. 2006. "Coccolith strontium to calcium ratios in *Emiliania huxleyi*: The dependence on seawater strontium and calcium concentrations." *Limnol. Oceanogr*. **51** (1), 310–320.

6. Mosekilde, L. 2000. "Age-related changes in bone mass, structure, and strength-effects of loading." *Z. Rheumatol.* **59** (Suppl 1), 1-9.

7. Sanchez, C., Arribart, H., and Guille, M. M. 2005. "Biomimetism and bioinspiration as tools for the design of innovative materials and systems." *Nat. Mater.* **4** (4), 277-288.

8. Sigel, A., Sigel, H. and Sigel, R., ed. 2008. *Biomineralization. From Nature to Application*. West Sussex: Wiley.

9. Weiner, S. and Dove, P.M. 2003. "An overview of biomineralization processes and the problem of the vital effect." *Rev. Mineral. Geochem.* **54,** 1-29.

10. Boskey, A.L. 2003. "Biomineralization: an overview" *Connect. Tissue Res*. **44** (Suppl. 1), 5-9.

11. Morse, J.W., Arvidson, R.S., and Lüttge, A. 2007. "Calcium carbonate formation and dissolution." *Chem. Rev.* **107** (2), 342-381.

12. Qin, C., Baba, O, and Butler, W. 2004. "Post-translational Modifications of SIBLING Proteins and Their Roles in Osteogenesis and Dentinogenesis." *Crit. Rev. Oral Biol. Med.* **15** (3), 126-136.

13. Salih, E., Wang, J., Mah, J., and Fluckiger, R. 2002. "Natural variation in the extent of phosphorylation of bone phosphoproteins as a function of *in vivo* new bone formation induced by demineralized bone matrix in soft tissue and bony environments." *Biochem. J.* **364** (Pt 2), 465-474.

14. Saad, F.A., Salih, E., and Glimcher, M.J. 2008. "Identification of osteopontin phosphorylation sites involved in bone remodeling and inhibition of pathological calcification." *J. Cell Biochem.* **103** (3), 852-856.

15. Salih, E. 2003. "*In vivo* and *in vitro* phosphorylation regions of bone sialoprotein." *Connect. Tissue. Res.* **44**, 223-229.

16. Salih, E. and Fluckiger, R. 2004. "Complete topographical distribution of both the *in vivo* and *in vitro* phosphorylation sites of bone sialoprotein and their biological implications." *J. Biol. Chem.* **279**, 19808–19815.

17. Choi, S., Kim, J., Kang, E., Lee, S., Park, M., Park, Y., and Lee, S. 2008. "Osteopontin might be involved in bone remodeling rather than in inflammation in ankylosing spondylitis." *Rheumatology (Oxford)* **47** (12), 1775-1779.

18. Reinholt, F., Hultenby, K., Oldberg, A., and Heinegard, D. 1990. "Osteopontin-a possible anchor of osteoclasts to bone." *Proc. Natl. Acad. Sci. U.S.A.* **87** (12), 4473–4475.

19. Ek-Rylander, B., Flores, M., Wendel, M., Heinegard, D., and Andersson, G. 1994. "Dephosphorylation of osteopontin and bone sialoprotein by osteoclastic tartrate-resistant acid phosphatase. Modulation of osteoclast adhesion *in vitro*." *J. Biol. Chem.* **269**, 14853-14856.

20. Katayama, Y., House, C., Udagawa, N., Kazama, J., McFarland, R., Martin, T., and Findlay, D. 1998. "Casein kinase 2 phosphorylation of recombinant rat osteopontin enhances adhesion of osteoclasts but not osteoblasts." *J. Cell Physiol.* **176**, 179-187.

21. Boskey, A.L., Maresca, M., Ullrich, W., Doty, S. B., Butler, W. T., and Prince, C.W. 1993. "Osteopontin-hydroxyapatite interactions *in vitro*: inhibition of hydroxyapatite formation and growth in a gelatin-gel." *Bone Miner.* **22**, 147-159.

22. Jono, S., Peinado, C., and Giachelli, C. 2000. "Phosphorylation of Osteopontin Is Required for Inhibition of Vascular Smooth Muscle Cell Calcification." *J. Biol. Chem.* **275**, 20197-20203.

23. Hunter, G. and Goldberg, H. 1994. "Modulation of crystal formation by bone phosphoproteins: role of glutamic acid-rich sequences in the nucleation of hydroxyapatite by bone sialoprotein." *Biochem. J.* **302**, 175–179.

24. Baht, G., O'Young, J., Borovina, A., Chen, H., Tye, C., Karttunen, M., Lajoie, G., Hunter, G., and Goldberg, H. 2010. "Phosphorylation of Ser136 is critical for potent bone sialoprotein-mediated nucleation of hydroxyapatite crystals." *Biochem. J.* **428**, 385-395.

25. Hunter, G. K., Hauschka, P.V., Poole, A.R., Rosenberg, L.C., and Goldberg, H.A. 1996. "Nucleation and inhibition of hydroxyapatite formation by mineralized tissue proteins." *Biochem. J.* **317**, 59-64.

26. Pampena, D.A., Robertson, K.A., Litvinova, O., Lajoie, G., Goldberg, H.A., and Hunter, G.K. 2004. "Inhibition of hydroxyapatite formation by osteopontin phosphopeptides." *Biochem. J.* **378** (Pt 3), 1083-1087.

27. Boskey, A.L., Spevak, L., Paschalis, E., Doty, S.B., and McKee, M.D. 2002. "Osteopontin deficiency increases mineral content and mineral crystallinity in mouse bone." *Calcif. Tissue Int.* **71** (2), 145-154.

28. Malaval, L., Wade-Guéye, N.M., Boudiffa, M., Fei, J., Zirngibl, R., Chen, F., Laroche, N., Roux, J.P., Burt-Pichat, B., Duboeuf, F., Boivin, G., Jurdic, P., Lafage-Proust, M.H., Amédée, J., Vico, L., Rossant, J., and Aubin, J.E. 2008. "Bone sialoprotein plays a functional role in bone formation and osteoclastogenesis." *J. Exp. Med*. **205** (5), 1145-1153.

29. Fujisawa, R., Butler, W.T., Brunn, J.C., Zhou, H.Y., and Kuboki, Y. 1993. "Differences in composition of cell-attachment sialoproteins between dentin and bone." *J. Dent. Res.***72** (8), 1222-1226.

30. Ritchie, H.H., Hou, H., Veis, A., and Butler, W.T. 1994. "Cloning and sequence determination of rat dentin sialoprotein, a novel dentin protein." *J. Biol. Chem.* **269**, 3698-3702.

31. Ritchie H.H., Wang, L.H., and Knudtson, K. 2001. "A novel rat 523 amino acid phosphophoryn: nucleotide sequence and genomic organization." *Biochem. Biophys. Acta.* **1520**, 212-222.

32. He, G., Ramachandran, A., Dahl, T., George, S., Schultz, D., Cookson, D., Veis, A., and George, A. 2005. "Phosphorylation of phosphophoryn is crucial for its function as a mediator of biomineralization." *J. Biol. Chem.* **280** (39), 33109–33114.

33. Saito, T., Arsenault, A., Yamauchi, M., Kuboki, Y., and Crenshaw, M. 1997. "Mineral induction by immobilized phosphoproteins." *Bone* **21**, 305-311.

34. van den Bos, T., Steinfort, J., and Beertsen, W. 1993. "Effect of bound phosphoproteins and other organic phosphates on alkaline phosphatase-induced mineralization of collagenous matrices *in vitro*." *Bone Miner* **23** (2), 81-93.

35. Qin, C., Brunn, J.C., Cook, R.G., Orkiszewski, R.S., Malone, J.P., Veis, A., and Butler, W.T. 2003. "Evidence for the proteolytic processing of dentin matrix protein 1: identification and characterization of processed fragments and cleavage sites." *J. Biol. Chem.* **278**, 34700-34708.

36. Sreenath, T., Thyagarajan, T., Hall, B., Longenecker, G., D'Souza, R., Hong, S., Wright, J.T., MacDougall, M., Sauk, J., and Kulkarni, A.B. 2003. "Dentin sialophosphoprotein knockout mouse teeth display widened predentin zone and develop defective dentin mineralization similar to human dentinogenesis imperfecta type III." *J. Biol. Chem*. **278** (27), 24874-24880.

37. Kim, J.W., Hu, J.C., Lee, J.I., Moon, S.K., Kim, Y.J., Jang, K.T., Lee, S.H., Kim, C.C., Hahn, S.H., and Simmer, J.P. 2005. "Mutational hot spot in the DSPP gene causing dentinogenesis imperfecta type II." *Hum. Genet.* **116** (3), 186-191.

38. Meenakshi, V.R, Hare, P.E., and Wilbur, K.M. 1971. "Amino acids of the organic matrix of neogastropod shells." *Comp. Biochem. Phys. B* **40** (4), 1037-1043.

39. Westbroek, P. and de Jong, E.W. 1982. *Biomineralization and Biological Metal Accumulation.* Dordrecht: D. Reidel Publishing Company.

40. Hecker, A., Testenière, O., Frèdèric, M., and Luquet, G. 2003. "Phosphorylation of serine residues is fundamental for the calcium-binding ability of Orchestin, a soluble matrix protein from crustacean calcium storage structures." *FEBS Letters*. **535** (3), 49-54.

41. Glazer, L., Shechter, A., Tom, M., Yudkovski Y., Weil, S., Aflalo, E.D., Pamuru, R.R., Khalaila, I., Bentov, S., Berman, A., and Sagi, A. 2010. "A protein involved in the assembly of an extracellular calcium storage matrix." *J. Biol. Chem*. **285** (17), 12831-12839.

42. Halloran, B., and Donachy, E. 1995. "Characterization of organic matrix macromolecules from the shells of the antarctic scallop, *Adamussium colbecki*." *Comp. Biochem Physiol.* **111** (2), 221-231.

43. Rusenko, K.W., Donachy, J.E., Wheeler, A.P., Suga, S., and Watabe, N. eds. 1991. "Purification and Characterization of a Shell Matrix Phosphoprotein from the American Oyster." *Surface Reactive Peptides and Polymers.* Tokyo: Springer-Verlag, 171–187.

44. Samata, T., Ikeda, D., Kajikawa, A., Sato, H., Nogawa, C., Yamada, D., Yamazaki, R., and Akiyama, T. 2008. "A novel phosphorylated glycoprotein in the shell matrix of the oyster *Crassostrea nippona*." *FEBS J.* **275** (11), 2977-2989.

45. Inoue, H., Ohira, T., and Nagasawa, H. 2007. "Significance of the N- and C-terminal regions of CAP-1, a cuticle calcification-associated peptide from the exoskeleton of the crayfish, for calcification." *Peptides* **28** (3), 566-573.

46. Alvares, K., Dixit S.N., Lux, E., Veis, A. 2009. "Echinoderm phosphorylated matrix proteins UTMP16 and UTMP19 have different functions in sea urchin tooth mineralization." *J. Biol. Chem.* **284** (38), 26149-26160.

47. Llies, M.R., Peeler M.T., Dechtiaruk A.M., and Ettensohn C.A. 2002. "Identification and developmental expression of new biomineralization proteins in the sea urchin *Strongylocentrotus purpuratus*." *Dev. Genes Evol.* **212**, 419-431.

48. Kröger, N., and Poulsen, N. 2008. "Diatoms-From Cell Wall Biogenesis to Nanotechnology." *Annu. Rev. Genet.* **42**, 83–107.

49. Schmid, A.M., Borowitzka, M.A. and Volcani, B.E. 1981. Morphogenesis and biochemistry of diatom cell walls. *Cytomorphogenesis in plants.* **8**, 63-97.

50. Bauerlein, E. 2003. "Biomineralization of unicellular organisms: an unusual membrane biochemistry for the production of inorganic nano-and microstructures." *Angew. Chem. Int. Ed. Engl*. **42**, 614–641.

51. Round, F. E., Crawford, R. M., and Mann, D. G. 1990. *The Diatoms: Biology and Morphology of the Genera.* Cambridge: Cambridge University Press.

52. Hildebrand, M., Doktycz, M. J. and Allison, D.P. 2008. "Application of AFM in understanding biomineral formation in diatoms." *Eur. J Physiol.* **456**, 127–137.

53. Losic, D., Pillar, R.J., Dilger, T., Mitchell, J.G., Voelcker, N.H. 2007. "Atomic force microscopy (AFM) characterisation of the porous silica nanostructure of two centric diatoms." *J. Porous Mater*. 14, 61–69.

54. Enache, M. and Potapova, M. 2009. "A new species of Sellaphora (Sellaphoraceae) from Hannaberry Lake, Arkansas, U.S.A." *Acta Bot. Croat.* **68** (2), 231–237.

55. Andreoli, C., Tolomio, C., Moro, I., Radice, M., Moschin, E. and Bellato, S. 1995. "Diatoms and dinoflagellates in Terra Nova Bay (Ross Sea-Antarctica) during austral summer 1990." *Polar Biol.* **15**, 465-475.

56. Gebeshuber, I.C. 2007. "Biotribology inspires new technologies." *Nano Today* **2** (5), 30-37.

57. Kilroy, C., Sabbe, K., Bergey, E.A., Vyerman, W. and Lowe, R. 2003. "New species of Fragilariforma (Bacillariophyceae) from New Zealand and Australia." *New Zeal. J. Bot.* **41,** 535-554.

58. Kröger N., Deutzmann, R. and Sumper, M. 1999. "Polycationic Peptides from Diatom Biosilica That Direct Silica Nanosphere Formation" *Science* **286**, 1129-1132.

59. Kröger, N., Lorenz, S., Brunner, E., and Sumper M. 2002. "Self-Assembly of Highly Phosphorylated Silaffins and Their Function in Biosilica Morphogenesis." *Science* **298**, 584–586.

60. Poulsen, N., Sumper, M., and Kröger, N. 2003. "Biosilica formation in diatoms: Characterization of native silaffin-2 and its role in silica morphogenesis." *Proc. Natl. Acad. Sci.U.S.A.* **100** (21), 12075–12080.

61. Poulsen, N., and Kröger, N. 2004. "Silica Morphogenesis by Alternative Processing of Silaffins in the Diatom *Thalassiosira pseudonana*."*J. Biol. Chem.* **279**, 42993–42999.

62. Wenzl, S., Hett, R., Richthammer, P., and Sumper, M. 2008. "Silacidins: Highly Acidic Phosphopeptides from Diatom Shells Assist in Silica Precipitation *In vitro*." *Angew. Chem. Int. Ed. Engl*. **47**, 1729–1732.

63. Kröger, N. and Sumper, M. 2004. "Silica formation in diatoms: the function of long-chain polyamines and silaffins." *J. Mater. Chem.* **14**, 2059-2065.

64. Kröger, N., Deutzmann, R., Bergsdorf, C., and Sumper, M. 2000. "Species-specific polyamines from diatoms control silica morphology." *Proc.Natl. Acad. Sci. USA* 97, 14133–14138.

65. Frigeri, L. G., Radabaugh, T. R., Haynes, P. A., and Hildebrand, M. 2006. "Identification of Proteins from a Cell Wall Fraction of the Diatom Thalassiosira pseudonana." *Mol. Cell. Proteomics* **5**, 182–193.

66. Sumper, M. and Brunner, E. 2006. "Learning from Diatoms: Nature's Tools for the Production of Nanostructured Silica." *Adv. Funct. Mater.* **16**, 17–26.

67. Safran, J.B., Butler, W.T., and Farach-Carson, M.C. 1998. "Modulation of Osteopontin Post-translational State by 1,25-$(OH)_2$-Vitamin D_3: Dependence on Ca^{2+} Influx. *J. Biol Chem*. **273** (45), 29935–29941.

68. Walsh, C. 2006. *Posttranslational Modification of Proteins: Expanding Nature's Inventory.* Englewood: Roberts and Company Publishers.

69. Puttick, J., Baker, E.N., and Delbaere, L.T. 2008. "Histidine phosphorylation in biological systems." *Biochim Biophys Acta.* **1784** (1), 100-105.

70. Minkin, C. 1982. “Bone acid phosphatase: tartrate-resistant acid phosphatase as a marker of osteoclast function.” *Calcif. Tissue Int.* **34** (3), 285-290.

71. Hayman, A.R., Jones, S.J., Boyde, A., Foster, D., Colledge, W.H., Carlton, M.B., Evans, M.J., and Cox, T.M. 1996. “Mice lacking tartrate-resistant acid phosphatase (Acp 5) have disrupted endochondral ossification and mild osteopetrosis.” *Development.* **122** (10), 3151-3162.

72. Ek-Rylander, B., and Andersson, G. 2010. “Osteoclast migration on phosphorylated osteopontin is regulated by endogenous tartrate-resistant acid phosphatase.” *Exp. Cell. Res.* **316** (3), 443-451.

73. Angel, N.Z., Walsh, N., Forwood, M.R., Ostrowski, M.C., Cassady, A.I., and Hume, D.A. 2000. “Transgenic mice overexpressing tartrate-resistant acid phosphatase exhibit an increased rate of bone turnover.” *J. Bone Miner. Res.* **15** (1), 103-110.

74. Veis, A., Sfeir, C., and Wu, C. B. 1997. “Phosphorylation of the proteins of the extracellular matrix of mineralized tissues by casein kinase-like activity.” *Crit. Rev. Oral Biol. Med.* **8** (4), 360–379.

75. Lasa, M., Chang, P. L., Prince, C. W., and Pinna, L. A. 1997. “Phosphorylation of osteopontin by Golgi apparatus casein kinase.” *Biochem. Biophys. Res. Commun.* **240** (3), 602–605.

76. Wu, C.B., Pelech, S.L., Veis, A. 1992. “The *in vitro* phosphorylation of the native rat incisor dentin phosphophoryns.” *J. Biol. Chem.* **267** (23), 16588-16594.

77. Redegeld, F. A., Caldwell, C. C., and Sitkovsky, M. V. 1999. “Ecto-protein kinases: ecto-domain phosphorylation as a novel target for pharmacological manipulation?” *Trends Pharmacol. Sci.* **20** (11), 453-459.

78. Zhu, X., Luo, C., Ferrier, J. M., and Sodek, J. 1997. “Evidence of ectokinase-mediated phosphorylation of osteopontin and bone sialoprotein by osteoblasts during bone formation *in vitro*.” *Biochem. J.* **323** (Pt 3), 637–643.

79. Hanks, S. and Hunter, T. 1995. “Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification.” *FASEB J.* **9**, 76-96.

80. Carling, D., and Hardie, D. G. 1989. “The substrate and sequence specificity of the AMP-activated protein kinase. Phosphorylation of glycogen synthase and phosphorylase kinase.” *Biochim. Biophys. Acta.* **1012**, 81-86.

81. Golsteyn, R.M., Mundt, K.E., Fry, A.M., and Nigg. E.A. 1995. “Cell cycle regulation of the activity and subcellular localization of Plk1, a human protein kinase implicated in mitotic spindle function.” *J. Cell Biol.* **129** (6), 1617-1628.

82. Zhang, C., and Zhang, R. 2006. “Matrix proteins in the outer shells of molluscs.” *Mar Biotechnol (NY).* **8**, 572-586.

83. Ishikawa, H. O., Takeuchi, H., Haltiwanger, R. S., and Irvine, K. D. 2008. “Four-jointed is a Golgi kinase that phosphorylates a subset of cadherin domains.” *Science* **321**, 401–404.

84. Frasson, M., Vitadello, M., Brunati, A. M., La Rocca, N., Tibaldi, E., Pinna,L. A., Gorza, L., and Donella-Deana, A. 2009. “Grp94 is Tyr-phosphorylated by Fyn in the lumen of the endoplasmic reticulum and translocates to Golgi in differentiating myoblasts.” *Biochim. Biophys. Acta* **1793**, 239–252.

85. Armbrust, E. V., Berges, J. A., Bowler, C., Green, B. R., Martinez, D., Putnam, N. H., Zhou, S., Allen, A. E., Apt, K. E., Bechner, M., Brzezinski, M. A., Chaal, B. K., Chiovitti, A., Davis, A. K., Demarest, M. S., Detter, J. C., Glavina, T., Goodstein, D., Hadi, M. Z., Hellsten, U., Hildebrand, M., Jenkins, B. D., Jurka, J., Kapitonov, V. V., Kröger, N., Lau, W. W., Lane, T. W., Larimer, F. W., Lippmeier, J. C., Lucas, S., Medina, M., Montsant, A.,Obornik, M., Parker, M. S., Palenik, B., Pazour, G. J., Richardson, P. M., Rynearson, T. A., Saito, M. A., Schwartz, D. C., Thamatrakoln, K., Valentin, K., Vardi, A., Wilkerson, F. P., and Rokhsar, D. S. 2004. “The Genome of the Diatom *Thalassiosira Pseudonana*: Ecology, Evolution, and Metabolism.” *Science* **306**, 79–86.

86. Bowler, C., Allen, A. E., Badger, J. H., Grimwood, J., Jabbari, K., Kuo, A., Maheswari, U., Martens, C., Maumus, F., Otillar, R. P., Rayko, E., Salamov, A., Vandepoele, K., Beszteri, B., Gruber, A., Heijde, M., Katinka, M., Mock, T., Valentin, K., Verret, F., Berges, J. A., Brownlee, C., Cadoret, J. P., Chiovitti, A., Choi, C. J., Coesel, S., De Martino, A., Detter, J. C., Durkin, C., Falciatore, A., Fournet, J., Haruta, M., Huysman, M. J., Jenkins, B. D., Jiroutova, K., Jorgensen, R. E., Joubert, Y., Kaplan, A., Kro¨ger, N., Kroth, P. G., La Roche, J., Lindquist, E., Lommer, M., Martin-Je´ze´quel, V., Lopez, P. J., Lucas, S., Mangogna, M., McGinnis, K., Medlin, L. K., Montsant, A., Oudot-Le Secq, M. P., Napoli, C., Obornik, M., Parker, M. S., Petit, J. L., Porcel, B. M., Poulsen, N., Robison, M., Rychlewski, L., Rynearson, T. A., Schmutz, J., Shapiro, H., Siaut, M., Stanley, M., Sussman, M. R., Taylor, A. R., Vardi, A., von Dassow, P., Vyverman, W., Willis, A., Wyrwicz, L. S.,Rokhsar, D. S., Weissenbach, J., Armbrust, E. V., Green, B. R., Van de Peer,Y., and Grigoriev, I. V. 2008. “The Phaeodactylum genome reveals the evolutionary history of diatom genomes.” *Nature* **456**, 239-244.

87. Poulsen, N., Chesley, P. M., and Kröger, N. 2006. “Molecular genetic manipulation of the diatom *Thalassiosira pseudonana* (Bacillariophyceae).” *J. Phycol.* **42**, 1059-1065.

88. Poulsen, N., Berne, C., Spain, J, and Kröger, N. 2007. "Silica immobilization of an enzyme via genetic engineering of the diatom *Thalassiosira pseudonana*." *Angew. Chem. Int. Ed.* **46**, 1843-1846.

89. Gordon, R., Sterrenburg, F.A.S., and Sandhage, K.H. 2005. A special issue on diatom nanotechnology. *J. Nanosci. Nanotech.* **5**, 1-4.

90. Röttger, S., White, J., Wandall, H.H., Olivo, J.C., Stark, A., Bennett, E.P., Whitehouse, C., Berger, E.G., Clausen, H., and Nilsson, T. 1998. "Localisation of three human polypeptide GalNAc-transferases in HeLa cells suggests initiation of O-linked glycosylation throughout the Golgi apparatus." *J. Cell Sci.* **111**, 45-60.

91. Mock, T., Samanta, M. P., Iverson, V., Berthiaume, C., Robison, M., M., Holtermann, K., Durkin, C., Bondurant, S. S., Richmond, K., Rodesch, M., Kallas, T., Huttlin, E. L., Cerrina, F., Sussman, M. R., and Armbrust, E. V. 2008. "Whole-genome expression profiling of the marine diatom Thalassiosira pseudonana identifies genes involved in silicon bioprocesses." *Proc. Natl. Acad. Sci. U.S.A.* **105**, 1579-1584.

92. Kröger, N., Dickerson, M. B., Ahmad, G., Cai, Y., Haluska, M. S., Sandhage, K. H., Poulsen, N., and Sheppard, V. C. 2006. "Bioenabled Synthesis of Rutile (TiO_2) at Ambient Temperature and Neutral pH." *Angew. Chem. Int. Ed. Engl*. **45**, 7239–7243.

93. Laemmli, U. 1970. "Cleavage of structural proteins during the assembly of the head of bacteriophage T4." *Nature* **227**, 680–685.

94. Smith, P. K., Krohn, R. I., Hermanson, G. T., Mallia, A. K., Gartner, F. H., Provenzano, M.D., Fujimoto, E.K., Goeke, N.M., Olson, B.J., Klenk, and D.C. 1985. Measurement of protein using bicinchoninic acid. *Anal. Biochem.* **150**, 76-85.

95. R. K. Iler,ed. 1979. *The Chemistry of Silica*. Wiley: New York.

96. Harrison, P.J., R.E. Waters and F.J.R. Taylor. 1980. "A broad spectrum artificial medium for coastal and open ocean phytoplankton." *J. Phycol*. **16**, 28-35.

97. Sheppard, V. Poulsen, N., and Kröger, N. 2010. Characterization of an Endoplasmic Reticulum-associated Silaffin Kinase from the Diatom *Thalassiosira pseudonana*. *J. Biol. Chem.* **285** (2), 1166-1176.

98. Hildebrand, M., Frigeri, L. G., and Davis, A. K. 2007. "Synchronized growth of *Thalassiosira pseudonana* (Bacillariophyceae) provides novel insights into cell wall synthesis processes in relation to the cell cycle." *J. Phycol.* **43**, 730-740.

99. Koresawa, M. and Okabe, T. 2004. "High-throughput screening with quantitation of ATP concentration: A universal non-radioactive, homogenous assay for protein kinase." *Assay Drug Dev. Technol*. **2**, 153–160.

100. Buss, J. E., and Stull, J. T. 1983. "Measurement of chemical phosphate in proteins." *Methods Enzymol.* **99**, 7–14.

101. Ringer, D. P. 1991. "Separation of phosphotyrosine, phosphoserine, and phosphothreonine by high-performance liquid chromatography." *Methods Enzymol.* **201 ,** 3–27.

102. Arnon, D. I. 1949. Copper enzymes in isolated chloroplasts. Polyphenoloxidase in *Beta vulgaris*. *Plant Physiol*. **24**, 1-15.

103. Packer, L., and Fleischer, S. 1997. *Biomembranes*. Academic Press Inc., San Diego, pp. 214–219.

104. Galbraith, D. W., Bohnert, H. J., and Bourque, D. P. 1995. *Methods in plant cell biology*. Academic Press Inc., San Diego, CA.

105. Brown, R.E., Jarvis, K.L., Hyland, K.J. 1989. "Protein measurement using bicinchoninic acid: elimination of interfering substances." *Anal Biochem.* **180** (1), 136-139.

106. Fini, C., Floridi, A., and Finelli, V. editors.1990. *Laboratory methodology in biochemistry: amino acid analysis and protein sequencing.* Boca Raton: CRC Press.

107. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.* **214**, 1-8.

108. De Castro, E., Sigrist, C., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P., Gasteiger, E., Bairoch, A., and Hulo, N. 2006. ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.* **34** (Web Server issue): W362–365.

109. Emanuelsson, O., Brunak, S., von Heijne, G., and Nielsen, H. 2007. Locating proteins in the cell using TargetP, SignalP, and related tools. *Nature Protocols* **2**, 953-971.

110. Elofsson, A., and von Heijne, G. 2007. "Membrane protein structure: prediction versus reality." *Annu. Rev. Biochem.* **76**, 125-140.

111. Diks, S.H., Parikh, K., van der Sijde, M., Joore, J., Ritsema, T., Peppelenbosch, M.P. 2007. *PLoS One* **2** (1), e777.

112. Waugh, D.S. Making the most of affinity tags. 2005. *Trends Biotechnol.* **23** (6), 316-320.

113. Mort, A. J. and Lamport, D. T. A. 1977. "Anhydrous hydrogen fluoride deglycosylates glycoproteins." *Anal. Biochem.* **82,** 289–309.

114. Wu, M.M., Llopis, J., Adams, S., McCaffery, J.M., Kulomaa, M.S., Machen, T.E., Moore, H.P., and Tsien, R.Y. 2000. "Organelle pH studies using targeted avidin and fluorescein-biotin." *Chem. Biol.* **7** (3), 197-209.

115. Kim, J.H., Johannes, L., Goud, B., Antony, C., Lingwood, C.A., Daneman, R., and Grinstein, S. 1998. "Non-invasive measurement of the pH of the endoplasmic reticulum at rest and during calcium release." *Proc. Natl. Acad. Sci. USA* **95**, 2997–3002.

116. Thuveson, M., and Fries, E. 2000. "The low pH in trans-Golgi triggers autocatalytic cleavage of pre-alpha -inhibitor heavy chain precursor." *J. Biol. Chem*. **275** (40), 30996-31000.

117. Vrieling, E.G., Gieskes, W. W. C., and Beelen, T. "Silicon deposition in diatoms: control by the pH inside the silicon deposition vesicle." *J. Phycol*. **35**, 548–559.

118. Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A., Lopez, R., Thompson, J. D., Gibson, T. J., and Higgins, D. G. 2007. "Clustal W and Clustal X version 2.0." *Bioinformatics* **23**, 2947–2948.

119. Mezghrani, A., Courageot, J., Mani, J.C., Pugniere, M., Bastiani, P., and Miquelis, R. 2000. "Identification of ERp29, an Endoplasmic Reticulum Lumenal Protein, as a New Member of the Thyroglobulin Folding Complex." *J. Biol. Chem*. **275** (3), 1920-1929.

120. Kroth, P. G. 2002. "Protein transport into secondary plastids and the evolution of primary and secondary plastids." *Int. Rev. Cytol.* **221**, 191–255.

121. Gething, M.J., and J. Sambrook. 1992. "Protein folding in the cell." *Nature* **355**, 33-45.

122. Gruber, A., Vugrinec, S., Hempel, F., Gould, S. B., Maier, U. G., and Kroth, P. G. 2007. "Protein targeting into complex diatom plastids: functional characterisation of a specific targeting motif." *Plant Mol. Biol.* **64**, 519–530.

123. Kilian, O., and Kroth, P. G. 2005. "Identification and characterization of a new conserved motif within the presequence of proteins targeted into complex diatom plastids." *Plant J.* **41**, 175–183.

124. Bhaya, D., and Grossman, A. R. 1991. “Characterization of gene clusters encoding the fucoxanthin chlorophyll proteins of the diatom Phaeodactylum tricornutum.” *Mol. Gen. Genet*. **229**, 400–404.

125. Lang, M., Apt, K. E., and Kroth, P. G. 1998. “Protein transport into "complex" diatom plastids utilizes two different targeting signals.” *J. Biol. Chem.* **273**, 30973–30978.

126. Apt, K. E., Zaslavkaia, L., Lippmeier, J. C., Lang, M., Kilian, O., Wetherbee, R., Grossman, A. R., and Kroth, P. G. 2002. “*In vivo* characterization of diatom multipartite plastid targeting signals.” *J. Cell Sci.* **115**, 4061–4069.

127. Arber, S., Krause, K. H., and Caroni, P. 1992. “s-cyclophilin is retained intracellularly via a unique COOH-terminal sequence and colocalizes with the calcium storage protein calreticulin.” *J. Cell Biol.* **116**, 113–125.

128. Monnat, J., Neuhaus, E. M., Pop, M. S., Ferrari, D. M., Kramer, B., and Soldati, T. 2000. “Identification of a novel saturable endoplasmic reticulum localization mechanism mediated by the C-terminus of a Dictyostelium protein disulfide isomerase.” *Mol. Biol. Cell* **11**, 3469–3484.

129. Anelli, T., Alessio, M., Bachi, A., Bergamelli, L., Bertoli, G., Camerini, S., Mezghrani, A., Ruffato, E., Simmen, T., and Sitia, R. 2003. “Thiol-mediated protein retention in the endoplasmic reticulum: the role of ERp44.” *EMBO J*. **22** (19), 5015-5022.

130. Blom, N., Sicheritz-Ponten, T., Gupta, R., Gammeltoft, S., and Brunak, S. 2004. “Prediction of post-translational glycosylation and phosphorylation of proteins from the amino acid sequence.” *Proteomics* **4** (6):1633-1649.

131. Rena, G., Bain, J., Elliott, M., Cohen, P. 2004. “D4476, a cell-permeant inhibitor of CK1, suppresses the site-specific phosphorylation and nuclear exclusion of FOXO1a.” *EMBO Rep*. **5** (1), 60-65.

132. Sarno S, de Moliner E, Ruzzene M, Pagano MA, Battistutta R, Bain J, Fabbro D, Schoepfer J, Elliott M, Furet P, Meggio F, Zanotti G, Pinna LA. 2003. “Biochemical and three-dimensional-structural study of the specific inhibition of protein kinase CK2 by [5-oxo-5,6-dihydroindolo-(1,2-a)quinazolin-7-yl]acetic acid (IQA).” *Biochem. J.* **374** (Pt 3), 639-646.

VITA

VONDA CHANTAL SHEPPARD

SHEPPARD was born in Ahokie, NC. She attended public school in Greenville, NC and received a B.S. in Biochemistry from the University of North Carolina at Greensboro in 2003 before coming to Georgia Tech to pursue a doctorate in Biochemistry.